Titre: Title:	Three essays on the economics of science policy: the impact of funding, collaboration and research chairs
Auteur: Author:	Seyed Reza Mirnezami
Date:	2015
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Mirnezami, S. R. (2015). Three essays on the economics of science policy: the impact of funding, collaboration and research chairs [Ph.D. thesis, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/2046/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/2046/
Directeurs de recherche: Advisors:	Catherine Beaudry
Programme: Program:	Génie industriel

UNIVERSITÉ DE MONTRÉAL

THREE ESSAYS ON THE ECONOMICS OF SCIENCE POLICY: THE IMPACT OF FUNDING, COLLABORATION AND RESEARCH CHAIRS

SEYED REZA MIRNEZAMI

DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION

DU DIPLÔME DE PHILOSOPHIAE DOCTOR

(GÉNIE INDUSTRIEL)

DÉCEMBRE 2015

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

THREE ESSAYS ON THE ECONOMICS OF SCIENCE POLICY: THE IMPACT OF FUNDING, COLLABORATION AND RESEARCH CHAIRS

présentée par : MIRNEZAMI Seyed Reza

en vue de l'obtention du diplôme de : Philosophiae Doctor

a été dûment acceptée par le jury d'examen constitué de :

M. JOANIS Marcelin, Ph. D., président

Mme BEAUDRY Catherine, Ph. D., membre et directrice de recherche

M. NIOSI Jorge E., Ph. D., membre

Mme YOUTIE Jan, Ph. D., membre externe

DEDICATION

To the Iranian Youth who gave their lives to defend their country against invasion (1980-1988)

ACKNOWLEDGEMENTS

I am very grateful to Prof. Catherine Beaudry for her patience, motivation, immense knowledge, and financial support. From her, I realized how much supportive a Ph.D. supervisor can be. Besides, I would like to thank Prof. Vincent Larivière for his help and support.

I would like to thank my family: my devoted wife for her endless love, my daughter for her smile at home, my parents for their support and encouragement throughout my life, and my parents in law and my sisters for their spiritual support.

RÉSUMÉ

Cette thèse étudie les déterminants qui influencent le nombre de citations, l'effet d'avoir une collaboration de recherche avec les scientifiques les mieux financés sur la productivité scientifique, et l'effet d'être titulaire d'une chaire de recherche sur la productivité scientifique. En supposant que le nombre de citations est une bonne mesure de l'impact de la recherche et, à son tour, d'un certain type de qualité, nous avons montré que le nombre d'articles et la visibilité d'un chercheur, le facteur d'impact de la revue, la taille de l'équipe de recherche, et le cadre institutionnel de l'université (effet fixe) sont les déterminants importants du nombre de citations. Cependant, nous avons constaté qu'il n'y a pas d'effet significatif du financement public ni du genre dans la plupart des domaines examinés.

Nous avons également développé un modèle théorique et proposé quelques hypothèses sur l'effet de la collaboration avec les scientifiques les mieux financés sur la productivité scientifique. Ce modèle a ensuite permis de valider les hypothèses à l'aide d'une analyse empirique et a montré que cette collaboration a un effet positif sur la productivité scientifique. Cet effet significatif peut exister à travers différents canaux: le transfert de connaissances tacites, davantage de publications scientifiques, des économies d'échelle dans la production de connaissances dues à de meilleurs équipements de recherche et un réseau de recherche élargi. Les résultats ont également vérifié l'effet positif du financement, l'effet positif du réseau (mesuré par le nombre de co-auteurs), l'effet en forme de U-inversé de l'âge, et le plus petit nombre de publications par les femmes par rapport aux hommes.

Enfin, nous avons fait une distinction entre les différents attributs des chaires de recherche et de leur effet sur la productivité scientifique. Une des questions importantes est de savoir si une chaire de recherche a encore une meilleure performance scientifique (par rapport aux non-titulaires) après avoir contrôlé par les fonds de recherche disponibles aux chercheurs. Pour étudier cela, nous avons utilisé une technique d'appariement pour identifier les paires de scientifiques (des titulaires et des non-titulaires de chaires) de même genre, financement et domaine de recherche. Après cette correspondance, nous avons constaté que l'effet du programme des chaires de recherche du Canada sur la productivité scientifique reste significatif et positif alors que l'effet des chaires industrielles et les titulaires de chaires nommés par les conseils canadiens subventionnaires fédéraux (CRSNG et IRSC) deviennent non significatif. Ce constat met en évidence l'efficacité de notre méthode de

technique d'appariement car avant l'appariement, tout type de chaire a un effet positif et significatif sur la productivité scientifique.

Ce constat met en évidence les attributs spéciaux du programme de chaires de recherche du Canada, qui sont différents des autres programmes de chaire. Ces attributs spécifiques peuvent pousser de manière significative la productivité scientifique. Entre autres, les chaires de recherche du Canada sont généralement associés à un certain degré de prestige et confèrent une plus grande visibilité pour recruter des étudiants talentueux ou pour développer une collaboration de recherche avec des scientifiques de haut niveau dans le domaine. Le fait que d'autres types de chaires de recherche, une fois appariés avec des scientifiques équivalents, n'ont pas d'impact sur la production scientifique en termes de quantité, ne signifie pas que ces titulaires de chaire sont des scientifiques de moindre envergure, mais qu'ils consacrent une partie de leur temps à d'autres efforts de nature plus pratique ou ayant un impact sociétal différent. Ainsi les universités maintiennent un équilibre entre la poursuite de la connaissance scientifique pure et son application à des avantages socio-économiques. En étudiant uniquement les articles scientifiques, il nous manque toutefois beaucoup d'information quant au rôle des professeurs d'université. Bien que non trivial, la recherche future devrait viser à ratisser plus large sur les réalisations, les résultats et les impacts de la recherche universitaire.

ABSTRACT

This thesis studies the determinants that influence the number of citations, the effect of having a research collaboration with top-funded scientists on scientific productivity, and the effect of holding a research chair on scientific productivity. Based on a review study by Bornmann and Daniel (2008), one can argue that non-scientific factors determining the decision to cite do not significantly alter the role of citation as a measure of research impact. Assuming that the number of citations is a good measure for research impact and, in turn, for a certain kind of quality, we showed that the number of articles and the visibility of a researcher, the impact factor of the journal, the size of the research team, and the institutional setting of the university are the important determinants of citation counts. However, we have found that there is no significant effect of public funding and gender in most of the domains examined. The point that funding amount is not a significant determinant of citation counts does not necessarily contradict the positive effect of funding on scientific productivity.

We also developed a theoretical model and proposed some hypotheses about the effect of collaboration with top-funded scientists on scientific productivity. We then validated the hypotheses with empirical analysis and showed that such collaboration has a positive effect on scientific productivity. This significant effect may exist through different channels: transfer of tacit knowledge, more scientific publications, economy of scale in knowledge production because of better research equipment, and expanded research network. The results also verified the positive effect of funding, the positive effect of networking (measured by number of co-authors), the inverted U-shaped effect of age, and the fewer number of publications by women compared to men.

Finally, we made a distinction between different attributes of research chairs and their effect on scientific productivity. One of the important questions is to find out whether a research chair still has better scientific productivity (compared to non-chair holders) after controlling for the research funds available to the researchers. To investigate that question, we employed a matching technique to identify pairs of scientists (chair and non-chair holders) of the same gender, funding and research field. After such matching, we found that the effect of the Canada research chair program on scientific productivity remains significant and positive, while the effect of industrial chairs and the chairs appointed by the Canadian federal granting councils (NSERC and CIHR) become non-significant. This finding highlights the effectiveness of our matching technique methodology;

because before matching, holding any type of chair had a positive and significant effect on scientific productivity.

This finding highlights the special attributes of the Canada research chair program, which are not replicated in other chairs. Those specific attributes may significantly push scientific productivity. For example, Canada research chairs are generally associated with some degree of prestige or higher visibility to recruit talented students or to have research collaboration with top scientists in the field. In addition, the Canada research chair program has a firm and efficient method of allocation (which is explained in the thesis). This approach institutionally synchronizes different chairs in universities and research fields. The fact that other types of research chairs, once matched with equivalent scientists, do not have an impact on scientific output in terms of quantity does not imply that these chair holders are lesser scientists, but that they are devoting part of their time to other endeavours of a more practical nature. Hence universities are maintaining a balance between the pursuit of pure scientific knowledge and its application to socioeconomic benefits. By solely studying scientific articles, we are missing a great deal of the university professors' activities. Although not trivial, future research should aim to cast a wider net on outputs, outcomes and impacts of university research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

The list of symbols and abbreviations presents the symbols and abbreviations used in the thesis or dissertation in alphabetical order, along with their meanings. Examples:

SP	Scientific Productivity
CRC	Canada Research Chair
SSHRC	Social Sciences and Humanities Research Council
CIHR	Canadian Institutes of Health Research
NSERC	Natural Sciences and Engineering Research Council
IV	Instrumental Variable
MT	Matching Technique
TFS	Top Funded Scientist

LIST OF APPENDICES

THESIS INTRODUCTION

This thesis investigates determinants of scientific productivity. It is estimated that 1.8 million articles are published each year, in about 28,000 journals. About half of the academic papers are read only by their authors and journal editors. It is also claimed that 90 percent of papers published are never cited (STM report, 2012). The number of citations, as a good proxy for research impact, and the number of papers published in peer reviewed journals are two factors by which scientists are assessed. Having known determinants of scientific productivity, governments and research grant awarding bodies can set some policies in order to increase the efficiency and effectiveness of this scientific productivity. Moreover, different research fields have distinctive institutional settings, and therefore research field-specific policies can help scientific communities to highlight their core competency.

Investigating science and technology policies should not be solely conducted based on empirical analysis; but the literature around theoretical modelling has not been well developed yet. Developing a theoretical model to predict scientific productivity and testing the prediction with real data is an important goal of this research. Inspired by Jensen and Meckling (1976) and by Durnev and Kim (2005), who developed concepts to model the incentives of firms' managers, we developed a theoretical model to show how the collaboration with top-funded scientists results in a higher number of peer reviewed journal articles. This model explains the cost–benefit analysis of a researcher to publish papers. Using contract theory and mechanism design literature, the model provides a theoretical prediction for an empirical investigation.

In terms of empirical analysis, different factors may affect research impact and scientific productivity of professors. Among other factors, funding amount, institutional support, gender, age, research field, and size of research team may impair or foster the scientific competency of faculty members, and hence lead to greater impact or greater quantity. The age effect may represent the life cycle trend in scientific productivity, indicating non-linearity or an inverted U-shape relationship between age and scientific productivity. The scientists' gender can be another determinant of scientific productivity, which may be explained by their administrative positions or marital status. Research funding for installing new infrastructure or equipment is a significant requirement for development and implementation of research. Collaborative research would also increase the knowledge diffusion and improve the research impact. Further, each scientific discipline has some field-specific characteristics that may determine scientific productivity.

Scientists benefit from institutional and government financial support, mainly paid from the taxpayer's pocket. Therefore, scientific productivity is a key concern with multiple stakeholders. In addition, countries' goals and preferences in science policy are not the same. This thesis focuses on the productivity of Canadian scientists to provide a better understanding for the case of Canada. There are also some studies explaining the trend of science policy in Canada and Quebec. According to Phillips (2010), Quebec has shown an acceptable productivity in terms of scientific productivity and patenting activities. Carew (2001) refers to the appropriate public investment in agricultural research and development in Canada, which was amplified by private funding since the 1990s due to budgetary considerations of public agencies. Niosi and Bas (2004) look at this trend differently and find that after the late 1980s the focus changed from agriculture to health products and services, which resulted in Canada's meeting the policy goals and being a leader in the field.

Furthermore, this research provides an appropriate understanding about the effect of age and gender on scientific productivity. This insight may help policymakers to develop some promotional scheme for scientists to encourage more publication with higher impact. The thesis also explores how much scientific productivity may depend on research field characteristics in the province of Quebec. This perspective may highlight the core competency of that province for knowledge production and may provide a better information for policymakers.

This thesis incorporates three papers about scientific productivity and research impact. To measure scientific productivity and research impact, we mainly focus on the number of articles and the number of citations respectively. However, measuring research impact is much more controversial. Chapter 1 reviews the related literature and chapter 2 frames the research questions and explains the methodology by which the hypotheses are being tested.

Chapter 3, 4, and 5 are three independent papers. The first paper, in chapter 3, investigates the determinants of citations (as a measure of research impact) by analysing information about researchers' funding and publication in the province of Quebec. This paper raises some novel and original research questions that have not been discussed previously. The first contribution of this paper is to analyze whether the various factors found in the literature are also positively associated, in our dataset, with changes in citation counts. Because those factors may be related to each other, the second contribution consists in collectively considering all of the mentioned factors in one econometrics model to test if their specific effects remain statistically significant. The study confirms the significant and positive relationship between the number of articles and citation counts. It also shows that scientists with more articles in higher impact factor journals generally receive more citations, as do scientists who publish with a larger team of authors. Moreover, male and female scientists receive the same number of citations, all else being equal.

The second and third papers, in chapters 4 and 5, analyse the number of articles as a measure of scientific productivity. The second paper tests whether collaboration with top-funded scientists increases the scientists' number of articles. Combining the known effect of networking and scientific collaboration, on the one hand, and scientist funding capacity, on the other hand, it is interesting to investigate the effect of collaborators' funding on scientific productivity. Based on a theory and empirical model developed in the second paper, the significance of collaboration with top-funded scientists in determining the number of scientific publications is shown.

In chapter 5 we focus on the effect of holding a "research chair" as a possible determinant of scientific publication. On the one hand, it may help the holder of the chair to be liberated from the constant quest for research funds or to have time to construct a more effective network, which may result in propelling future knowledge production. On the other hand, greater scientific productivity may simply be the effect of the past productivity of a scientist, implying an intrinsic ability of scientists in conducting research and/or in mobilising effectively their extensive networking capacity. According to the chair characteristics, the networking and prestige effect of "holding a research chair" may be mixed with the effect of funding. To address this issue, we use a matching technique and compare two "chaired" and "non-chaired" scientists who have similar funding and the same gender, and who work in the same research field as each other. Following the methodology employed by Bérubé and Mohnen (2009), the selection is made by generating propensity scores and choosing the pairs of scientists with the closest scores to each other. The new data set consists of "twin" scientists who are similar to each other in terms of funding, gender, and research fields. This technique improves the robustness and reliability of the result because it disentangles the prestige effect from the funding effect of the chair. In such a setting, "holding a research chair" becomes a better and more informative signal for the prestige of scientists.

Chapter 1 **LITERATURE REVIEW**

1.1 Introduction

This thesis identifies main determinants of scientific productivity and research impact. To measure scientific productivity, we mainly focus on the number of articles. However, measuring research impact is much more controversial. The sub-section 1.2 reviews the literature about the main determinants of scientific productivity and research impact. Some investigations in the literature identify age, gender, public grants and private contracts, research field, institutional characteristics, and some other factors as common determinants of scientific productivity and research impact. In none of mentioned factors, there is a sort of consensus. Studies have been conducted with different research methodologies using unrelated data sets. It is not un-expected to have distinctive conclusion and analysis. The following literature review gives the opportunity to identify the plausible determinants of research impact and scientific productivity. We try to have the best possible dataset and the most appropriate specification in empirical analysis in order to have a robust analysis. This implies having no assumption and pre-judgment on the effect of factors. However, where possible, we investigate whether the context of one studies may significantly affect the result of its analysis. Different points of view toward using number of citations as a measure of research impact has been discussed in sub-section 1.3.

1.2 Determinants of scientific productivity and research impact

The following determinants mainly refer to the required skills and capabilities for generating scientific outcomes in general, which are mainly reflected in the form of peer-reviewed publication in scientific journals.

1.2.1 Age

Kyvik and Olsen (2008) review different point of view in literature explaining the age effect on academic productivity measured by the number of articles and categorize them in six groups. All of the following hypotheses have been verified based on the data set used in different studies: (a) The utility maximizing hypothesis implies that academic staff conduct less research as they age because the expected utility of time spent on research diminishes; (b) The seniority burden hypothesis refers to the increasing administrative load as a career advances, which decrease the focus on scientific endeavours; (c) The cumulative disadvantage hypothesis suggests that scientists who do not win research awards will gradually lose their incentives for further research; (d) The age decrement hypothesis proposes that older scientists mostly conduct research with a lower intellectual and physical level than that of their younger colleagues; (e) The obsolescence hypothesis implies that the younger scientists use novel tools, techniques, and methodologies for research more easily than their older colleagues; (f) The intellectual deadlock hypothesis suggests that older scientists have less tendency to "reorient their research towards new scientific or social problems" (Kyvik and Olsen, 2008, pp. 442).

In addition to the above hypotheses, two trends regarding the effect of age can be observed in empirical investigations. First, some articles assess the life cycle trend in economic activity, referring to the non-linearity of human productivity during life (Becker (1962) is an example for this group). A number of articles show that productivity follows an inverted U-shape relationship with age, which can be justified by the optimization of the trade-off between the cost of human capital investment at a younger age and its return as benefit at an older age. For instance, Bernier et al. (1975) show an increase in impact and quantity of publication until the age of 44 and a diminution thereafter.

Some evidence is also found about the quadratic effect of age on scientific productivity. Turner and Mairesse (2005) investigate 500 French physicists for period of 1980-1997 and find linear and quadratic variation for the number of publications and for the number of citations with increasing age, controlling individual and environmental characteristics. Similar findings are observed by Gonzalez-Brambila and Veloso (2007), as they find a quadratic effect of age. In that paper, however, no strong evidence of investment motivated behaviour[1] was found: "researchers at 65 years are as productive as those at 43, and more than when they finished their PhD" (pp. 1050). They also argue that it may be due to the Mexican system which provides enough incentives for scientific production regardless of age.

[1] This behaviour explains that people use the resource of "age" to produce leisure and maximize it over the life cycle. In principle, because there is a trade-off between work-time and leisure, economic maximization models generally result in a peak in work-time.

The second group of articles generally find that scientists' academic productivity decreases as they age. Diamond (1986) finds that both impact (measured by the number of citations) and quantity of scientific publication (measured both in terms of the number of articles and number of pages) decrease with respect to age. In another interesting study, Levin and Stephan (1991) find that scientist productivity decreases with respect to age in five of the six disciplines[2] examined, controlling for all other probable effects. In addition, from the result of theoretical modelling, they attribute this phenomenon to the vintage effect[3]. In another investigation at the institutional level, Bonaccorsi and Daraio (2003) examine the influence of age distribution of researchers on scientific productivity of the Italian National Research Council and find that scientific productivity declines when the average age of researchers increases.

Other arguments indicate that the age effect on scientific productivity varies between different disciplines (Kyvik, 1990). The author indicates that the productivity level in the social sciences is independent of age. In the humanities, however, the publishing activity declines during the age-period of 55-59, but is followed by a new peak in the group of 60 years old and over. There is a different story in the medical sciences in which the productivity falls when researchers reach 55 and older. In the natural sciences, productivity continuously decreases with aging. Such differences come from the different nature of disciplines. As new scientific tools, methods, and equipment are continuously introduced, older researchers may have problems becoming familiar with them, and this generally referred to as the obsolescent effect in the literature (Kyvik, 1990). Vice versa, aging has less of an effect when the knowledge production has a lower speed, such as in the social sciences and humanities.

1.2.2 Gender

Gender effect is known as a significant determinant of scientific productivity in the literature. Hesli and Lee (2011)find that, in the field of political science, women publish less than what men do. Long (1992) obtains slightly different results by looking at the scientific productivity of

---

[2] Oceanography, geophysics, geology, solid-state/condensed-matter physics, atomic and molecular physics are the five disciplines and particle physics is the sixth one.

3 The vintage effect supposes that new capital (here read knowledge) is more valuable than old capital because new capital is produced via new tools, techniques, and methodology.

biochemists and testing three dimensions of productivity: publication counts, level of contribution in a team-work, and research impact (number of citations and the impact of the journal). In terms of publication counts, he shows that women are less productive in the first decade of their career but are more productive afterwards. In the other two dimensions, there is no difference between men and women in terms of the level of contribution and citation.

In terms of citations, Aksnes et al. (2011) show the female publications are significantly a bit less cited than are the male's. This is due to the cumulative advantage effect, referring to the fact that the marginal increase in citation is growing with an increase in publication output and because men have more publications, they can benefit more from this advantage, and hence have more number of citations for their papers.

Some researchers argue that differences in research productivity between men and women come from administrative positions of researchers and their marital status. For example Xie and Shauman (1998) conclude that gender differences in research productivity has declined over time, while at the same time the population of female scientists has proportionally increased (this decline is also observed in Abramo et al. (2009a)). Xie and Shauman (1998) also indicate that "women scientists publish fewer papers than men because women are less likely than men to have the personal characteristics, structural positions, and facilitating resources that are conducive to publication" (pp. 863).

Fox (2005) argues that the effect of gender is complex in a way that it is not possible to simply separate the effect of married and single researchers. He also refers to the career of spouse and family composition as two important factors of such complexity. For example, women with preschool children show higher productivity than women without children and women with school-age children. Nakhaie (2002) also investigate the gender effect in different durations of time and argues that Canadian female professors publish less than their male colleagues do, both in a lifetime period and during a shorter period, but such effect is higher for the longer period.

In addition to the papers which demonstrate the differences between men and women, some papers try to justify such differences. For example, Kyvik and Teigen (1996) develop a multiple-regression model to find determinants of individual productivity differences between men and women. Using data gathered by questionnaire in the spring of 1992, they find that childcare and lack of research collaboration are the two factors which negatively affect the scientific publishing of women and that such differences vary among nations due to factors such as the social security system or the child education system. In another study, Leahey (2006a) argues that the reason of low productivity of women is that women specialize less than men, which is an important factor for research productivity.

1.2.3 Private funding

Reviewing some socio-demographic factors, we now turn to two main funding aspect of scientific production. Balconi and Laboranti (2006) argue that in applied fields such as micro-electronics, there are two stages for the establishment of a new technology: first, the discovery of new avenues for potential commercial activities and second, the realisation of these possibilities by developing new industrial products via answering to the specific research question. To show that private funding is essential in some fields, the paper shows that public funding is suitable for explorative and uncertain research in terms of economic return (i.e. the first stage) and it is only the industrial support that can facilitate reaching the second stage.

Manjarrés-Henríquez et al. (2009) find that R&D contracts with industry are effective only if the professor is in need of money. In the other words, “R&D contracts with industry and academic research activities have synergistic effects on scientific production, but only when R&D contracts account for a small percentage of a researcher’s total funding, otherwise, there are decreasing marginal returns to scientific output” (pp. 799).

Geuna and Nesta (2003) claim that increased industrial funding will force researchers to shift to more applied research, neglecting their normative responsibilities for knowledge development. Similarly, Partha and David (1994) argue that institutions and social norms of open science are functionally maximizing the long term growth of scientific knowledge, but are not powerful to be socially optimum in terms of producing economic rents and commercial outputs from the existing stock of scientific knowledge. However, they explain that rules and reward systems in both university and industry symbiotically evolve to develop both science and technology.

Similarly, Kleinman and Vallas (2001) address a paradox called “the industrialization of the academy and the collegialization of industrial research” (pp. 451). In other words, scientists and engineers in private industry get a higher level of autonomy and control while scientists working in university settings are faced with the opposite case. In their empirical analysis, they address this paradox by arguing that norms, codes and practices in industry are penetrating the academy and vice versa. However, such infiltrating is asymmetric in favour of industrial norms.

From another point of view, Tijssen (2004) explains the declining trend in output of corporate research articles as the new orientation towards commercial application of in-house research results. The industry side increasingly tries not to publicize its research results. Goldfarb (2008) refers to the same story in university and finds that having collaboration with ‘research sponsor’ results in a fall in the number of publications by 25%. In another study of the engineering faculty in Germany, Hottenrott and Thorwarth (2011) indicate that the share of private funding in universities is high enough[4] to negatively affect the scientific publications.

In terms of private funding, Gulbrandsen and Smeby (2005) find that industry funding is correlated with “new and interesting research topics and is prerequisite to accomplishing expensive and interesting projects” (pp. 947), without any significant negative influence on researcher’s autonomy. Furthermore, such projects and research ideas generally result in more patents, commercial products, development of spin-off companies, and involvement in consulting work. Although private funding is correlated with both scientific and commercial products, they are not directly correlated with each other. Their data also show that professors with industrial funding collaborate more with other researchers both in academia and in industry.

In terms of the effect of such funding on industries, Berman (1990) finds that privately funded research has also speeded up the “transfer and utilization of academic research in industry” (pp. 353), via developing joint projects, hiring students, and faculty consulting.

Although the private funding is mainly provided by companies in the private sector, there are some examples of such funding in the form of philanthropic grants, which are provided by philanthropists.

[4] The paper finds that in the low share of private funding of total funding, the effect is not necessarily negative.

1.2.4 Public funding

Pavitt (2001) refers to the importance of public support for scientific infrastructure development and highlights its role in the effectiveness of public grants in the US. In another study, Pavitt (2000) argues that equipment and networks are necessary for the development and implementation of research, and is thus costly. Looking at the allocation of infrastructure grants in Australia and Britain, Harman (2000) identifies research strengths, maintaining research impact, and distributing research funds selectively with considerable administrative discretion as the important criteria for such allocation.

Providing required funding for conducting research via different science policies has an important role to push scientists toward the production of more effective and efficient research. Crespi and Geuna (2008) find that there is a delay for such policies to be effective. They find relatively little evidence of a positive impact of science policy in short lags (2-3 years), but the ‘total cumulated effects’ are significant for publication counts with a 5-year lag and for citations with a 6-year lag. Similarly, Ravenscraft and Scherer (1982) find an average lag of 4-6 years for the effectiveness of industrial research and development policies.

One well accepted justification for developing public support of science is that it is public good (Partha and David, 1994). Callon (1994) further argues that science is not a public good because of its intrinsic and natural properties[5], but a public good due to the fact that it is a source of diversity and flexibility. In addition, Crespi and Geuna (2008) advocate the public good nature of science by showing that science policy not only has a direct effect on one country and that there is a significant spillover between countries but also that spillover from the US to other countries are higher than in the reverse direction.

Pavitt (1991) argues that high attention to the public good justification may result in policies with two main unwanted implications: (1) policies of high priority for basic research disregard the benefits from training and unplanned discoveries; and (2) policies considering economies of scale with much focus on larger units in basic research are not necessarily working efficiently.

---

[5] He means that it is ***possible*** to make science as private good.

It is not possible to claim for a uniform argument about the effect of public funding. Appleyard (1996) indicates that public grants may have different effects in two countries or in two industries. The paper compares the semiconductor sector with the steel industry and finds greater effect of public recourses in the former one. She also finds that the public funding in universities plays a more important role in Japan than in the United States. Differences between countries are also investigated for Europe by Geuna (2001). For example, he mentions that some countries such as the UK pursue mission-oriented plans for some specific sectors and follow selective policies for some particular technologies, while in countries such as Italy, distributive and proportional policies are executed, which cover most of sectors and technologies.

Beise and Stahl (1999) show the positive effect of publicly funded research in universities on developing new products and processes in Germany. In a review article, Salter and Martin (2001) compare the papers investigating the economic effects of publicly funded basic research based on the methodology used in the papers. On the one hand, most of the econometric studies say that the economic benefits of publicly funded research in universities are crucial and substantial. On the other hand, the research based on surveys and case studies provide results, varying in scientific field, technology and industrial sector. The authors suggest the following six types of contribution of publicly funded research: source of new useful knowledge, new instrumentation and methodologies, skills developed by those involved in carrying out basic research, expansion of national and international networks, dealing with complex problems[6], and creation of spin-off companies.

As discussed in Salter and Martin (2001), there are two distinct forms of justification for public grant awarding bodies to award financial support. First, some justifications rely on lack of instruments and laboratory tools for conducting research and most such needs can be met by awarding infrastructure grants. The second group of justification focuses on the need for hiring new researchers/students or making/expanding of scientific network, which needs financial resources and the public grants cover them.

[6] The paper argues that solving the complex problem provides great benefit for the firms and organizations facing such problems.

Public funding can also result in knowledge flows to the private sector. Jaffe (1989) finds that university research significantly affects the number of patents in the area of drugs, medical technology, electronics, optics, and nuclear technology. Narin et al. (1997) have also provided some evidence of a positive effect of US public research on patent by tracing the number of papers (written by scientists in academy, government and other public institutions) cited in patents.

The fall in public funding research has some influence on the university research trend. Geuna (2001) argues that the fading role of public funding can result in over-use of resources, focus on short-term research endeavor, conflict in incentive structures, and “exacerbation of the impact of cumulative and self-reinforcement phenomena present in the process of scientific production” (pp. 626).

In terms of the purpose of funding, two different objectives can be identified: funding for operational cost and funding for buying infrastructure. For example, the Canadian Foundation for Innovation (CFI) is designed to meet “the current needs of the research community, its partners and stakeholders. The CFI funding architecture covers the full spectrum of infrastructure: projects to attract a leading researcher; team-led innovative projects that have a structuring effect for an institution or a region; and large-scale national projects”. Generally speaking, the Canadian government pursues research funding through three different areas of health science, social science & humanities, as well as engineering and natural science[7].

1.2.5 Collaborative research

Using data on collaborative research conducted in Canada, Godin and Gingras (2000) highlight the positive effect of collaborative research on scientific publication. Johnes (1988) indicate that the larger number of staff at a department and the larger number of co-authors of articles can significantly explain the higher research productivity in UK economics departments. Scientific collaboration, however, depends on the academic field. Abramo et al. (2009b) show that in the interdisciplinary fields there are more collaborations than in intramural fields and also that collaboration with foreign organizations are more in the basic fields than in the applied fields. In

[7] http://www.innovation.ca/en/OurFunds

addition, Ynalvez and Shrum (2011) indicate that in a resource-constrained developing country, scientific collaboration does not significantly affect research productivity, although most its scientists are involved in research collaboration even despite the coordination difficulties. However, such collaboration can result in professional opportunities and scientific rewards for researchers.

Networking is an important concept in the process of knowledge production. Not only are the scientists interested in networking, but people in industry also benefit from it both via patenting and publishing. Hicks (1995) explains reasons why industries publish scientific papers as a necessity to give a signal for the presence of tacit knowledge in the firms and then to make new linkages with the other organizations by barter-governed exchange of scientific and technical knowledge. The publications bring more reputation for the firms, promoting the firms in the knowledge exchange.

Chwe (2000) analyses social network using a game theory model and identifies three levels of hierarchy as the minimal sufficient structure of network for efficient collaboration. These hierarchical stages are 'initial adopters', 'followers', and then 'late adopters'. The paper indicates that "a communication network helps coordination in exactly two ways: by informing each stage about earlier stages, and by creating common knowledge within each stage" (pp. 1). It seems that the implication of such structure in the scientific community would be a network where the scientists collaborate and produce new productions, knowing the past information.

Most of the studies on the effects of network rely on co-authorship as a proxy of scientific collaboration. Melin and Persson (1996) argue that some collaboration does not generate papers and some co-authored articles should not necessarily be counted as the result of real scientific collaboration. Moreover, Katz and Martin (1997) argue that co-authorship is just a partial indicator of collaboration. They also mention that the concept of 'research collaboration' is hard to uniquely define: First, it is a social relationship and it is not possible to identify the beginning and the end; second, determinants of scientific collaboration depend on institutions, fields, sectors, countries, and time. As a result, an accurate and appropriate definition of collaboration should be altered to match the context of study.

There are also some concerns about the measurement of scientific collaboration between institutions in different countries (international collaboration). For example, Luukkonen et al. (1993) point out the difference between absolute and relative measurements. They argue that the normalized/relative measures reflect the differences in country size and the intensity of collaborative links, while the absolute measures reveal the centrality and importance of each country in the science network. Using such a distinction, it is possible to more efficiently consider the effect of international collaboration on the scientific productivity.

1.2.6 Research context and environment

Receiving support from environment would be an important factor in promoting scientific productivity and research impact. On top of that, each scientific discipline have some field-specific characteristics that may determine scientific productivity. Looking at nanotechnology and human genetics in Europe and the US, Heinze et al. (2009) show that small size of research group, sufficient access to the various technical skills, and an appropriate leadership result in an improvement in research creativity. Using laboratory level evidence from a large European university in the period of 1993–2000, Carayol and Matt (2004) analyze the typology and organization of research laboratories and find a correlation between individual productivity and the human resource structure of laboratories. The paper indicates that a combination of full-time researchers and university professors significantly results in high scientific productivity. It shows that university professor provides mentorship/supervision and funding but fruitful research activities require effective collaboration with full-time research staff.

Although most of Universities in Canada are publicly funded, university type and university governance may be another factor determining scientific productivity. Jordan et al. (1989) raise the question of efficiency of private universities in terms of research and teaching. They find strong and significant evidence that private institutions have a larger average productivity. Further investigating the question, Golden and Carstensen (1992) find that there is no difference between public and private universities in terms of research productivity when controlling for research support from leading research foundations and department faculty rating. They point out that the very nature of public and private institutions are different in a way that public institutions have more of a teaching and service role than private ones. They argue that public/private nature of an institution does not have an effect on scientific production per se, hence contradicting the result of Jordan et al. (1989). There may be some career related factors. Manjarrés-Henríquez et al. (2008) find that administrative positions within the university positively affect scientific productivity especially when a researcher is involved in both research and university-industry relationship. Such positions are important because their holders are the decision makers about policies of universities. Similar results are provided by Carayol and Matt (2006).

In another study by Jungnickel and Creswell (1994) the author indicate that the departmental workplace is very important for the scientist to generate research products from available resources. The factors that correlate with departmental workplace and that have significant effect on scholarly productivity are ‘off-campus conversation with the colleagues about research activities’ and ‘amount of time in research compared with other work-load’. Some articles refer to department characteristics as important factors in scientific production. Buchmueller et al. (1999) indicate that graduate school faculty size is a significant determinant of the research proficiency of graduates. Jordan et al. (1988, 1989) indicate that research productivity is positively associated with department size but that this effect becomes weaker as the size increases. Smeby and Try (2005) indicate that department climate (based on a qualitative questionnaire), age structure, and the proportion of faculty members with PhD degree affect the number of papers published by faculty members. Although this interesting factor can explain the scientific productivity but we do not have it in our database.

To justify the differences between disciplines, Baird (1986) shows that a large research laboratory in chemistry, scholarly apprenticeship approach in history, and research over practice in psychology are important factors in scientists’ productivity. In another comprehensive study, Baird (1991) refers to the productivity and citation pattern differences among disciplines.

Regarding cultural differences, Handberg (1986) make a comparison between the academic productivity in western and non-western society and indicates that physical and intellectual isolation in addition to the relative lack of significant meetings, conference, symposium, or convention of their disciplines or research areas negatively affect research in non-western society. Second, in non-western academic community the collaborative work may fail because the scientist are literally the only specialist in a particular topic and there are less competent interested graduate students or colleagues to discuss theoretical and experimental problem. Third, teaching is the primary reason that these institutions exist, which results in high teaching load and hence it hinders research activities.

Lertputtarak (2008) investigates the case of Thailand and finds that research productivity mainly correlates with desirable factors, which influences the willingness of scientists for conducting research and increases the individual's motivation. The author refers to the following factors: (1) to reduce teaching work-loads; (2) to overcome the inequity in salary; (3) to provide sufficient research facilities, consisting of resources, materials, machinery, equipment, research assistants, technicians, facilitators, databases, books, and stationery; (4) to set appropriate financial regulation and policies; (5) to top up the insufficient research funding; (6) to form positive academic and provide "both socializing and reinforcing organizational messages about norms, values and expectations concerning research" (pp. 205). We use university dummy variable to control the research characteristics and environment.

1.2.7 Prestige

There are generally two types of prestige. Regarding the effect of university prestige on research productivity, Crane (1965) finds that scientists at major universities are more likely to be productive and to get academic ranking promotion than scientists at minor universities. The main justification for such phenomenon comes from the fact that more prestigious universities are better able to select the best students, are richer and hence can recruit better researchers, and are well-endowed to provide research procurement. In a similar study, Turner (2005) indicates that graduates from the "French Grandes Ecoles" publish more than other French graduates. Groot and Garcia-Valderrama (2006) also show that larger research groups generate better research impact.

Focusing on the role of university prestige in academic productivity, Long et al. (1979) find a positive and significant correlation between the prestige of the alma matter of a scientist and where he is employed subsequently. They also indicate that graduating from a prestigious university has positive effect on citations but not on the publication counts. Furthermore, Zhou et al. (2012) look at the concept of prestige from another perspective. They find that the papers which are cited by prestigious scientists, regardless of the number of citation, have better research impact than papers which are cited by 'ordinary' scientists. This finding indirectly implies that scientist prestige can be an appropriate proxy for the generation of high impact research.

The individual prestige can be measured by some titles and awards. In Canada, research chair system is a set of programs to award title of research chair to different scientists along with some special financial support. In Canada, there are three types of research chair: (1) research chairs which are awarded by industry and generally referred to as industrial chairs; (2) research chairs which are awarded by Canadian federal funding agencies such as the Natural Science and Engineering Research Council (NSERC) and the Canadian Institutes of Health Research (CIHR); and (3) the 'Canada research chairs', whose holders are assumed to already achieve research excellence in one of the main fields of research.

A number of authors tried to highlight the functions and characteristics of research chairs. Cantu et al. (2009) showed the research chair program would be a good strategy for implementing knowledge-based development. In study on German universities, Schimank (2005) argued that chair-holders are small businessmen with high job security and no bankruptcy in addition to a good level of freedom of teaching and research, indicating that research chairs have the characteristics of job security and sovereignty. Individual prestige can be measured by dummy variables of research chair.

1.2.8 Initial condition of researchers

Allison and Stewart (1974) use the concept of cumulative advantage to show how the difference between two researchers in their early career can be amplified over time. Similar to a feedback loop phenomenon, highly productive scientists maintain or increase their productivity, while less productive scientists produce even less later on, implying that the spectrum of research productivity becomes wider over time. This finding is verified in both cross sectional (for different disciplines) and longitudinal analyses (for cohorts) of publications and citations. In a similar study, Allison et al. (1982) use true cohorts of biochemists and chemists, and show that the inequality increases over time for publications but not for citations per se. However, inequality for citations to recent publications has increased but this is mainly due to increased inequality of publications, which does not come from exclusive cumulative advantage of citations. We use year dummy to control the effect of each year. This may partially explain the condition of researcher affected by each year characteristics.

1.3 Measuring research impact and scientific productivity

Measuring quality of research is a controversial topic in literature. Most of the mentioned factors identifying scientific productivity can be also considered as determinants of research impact and number of citations. However, there are also some exclusive discussion about research impact in the literature. Ranking of journal and researcher's affiliation would be a proxy for research impact. Frey and Rost (2010) compare three types of ranking based on the number of articles, number of citations, and membership of editorial board and of academic associations. The paper indicates that these rakings are not compatible with each other and suggests the use of multiple measurements instead of citation or publication counts. Van Raan (2005) criticize the applicability of university rankings such as the Shanghai ranking for evaluating academic excellence by noting that the 'affiliation', as an important factor reflecting research atmosphere, is not well addressed in those ranking.

Meho (2007) argues that a very basic analysis of citations can indicate how much an article is used in the scientific community. Aksnes (2006) shows that citation counts match with own assessments of authors about their scientific contribution. However, the paper implies that citation counts are not reliable at the level of the individual article and average citation rate of the subfield is better indicator than the journal's one. Citations are nevertheless a useful raw material to assess the ranking of scientist (Radicchi et al., 2009).

Along these lines, Kostoff (1998) investigates the theory of citation and suggests that every citation results from the combination of two main reasons: the real component of intellectual heritage and random components of self-interest. Although there is a random component, the author argues that the random effect disappears in the aggregation of citation and therefore the number of citations is a good indicator of the "impact" of research. Phelan (1999) provides the same justification.

However, Amsterdamska and Leydesdorff (1989) refer to some cautionary about the use of citations. The authors argue that a cited article is used to either make a linkage with the current literature or to be used as an evidence to complete the justification. Therefore, a cited paper may be used in the citing papers with different purposes and with different implications. Furthermore, they argue that the citation analysis is just one aspect of knowledge production process.

Moed et al. (1985) also note that citations just refer to impact and are not a good proxy for research quality. Any publication should nevertheless have a minimum quality to generate an impact. For example, the paper mentioned the visibility of journals and the extent to which researchers provide a public service as two important determinants of citing a particular paper that do not have strong correlation with research quality. Going further from the recent argument, Olson et al. (2002) claim that even publishing does not have a two-way correspondence with minimum research quality because for example in Medicine there is a bias for publishing research with positive results.

It worth to mention that publication count without citation effect does not have anything about research impact, according to (Butler, 2003a); Butler (2003b). Nonetheless, in addition to these quantitative indicators, Butler (2007) argues that some qualitative assessment, like peer review, should be performed and combined with metrics to provide a fair and efficient overview of research impact. Fisher and Powers (2004) also note that peer review can indeed improve research impact. From another causal direction, Groot and Garcia-Valderrama (2006)find that the quality of peer review assessment in the Netherlands can be partially explained by number of publications in top-class and excellent international refereed journals, by the number of international proceedings, and by the number of Dutch journal articles.

The impact factor of the journal may be one candidate to measure research impact or prestige. However Seglen (1997b) notes that it cannot be an appropriate measure for research impact. The paper argues that citation rates determine the journal impact factor but nothing can be said about the reverse. In addition, journal impact factors in basic research are naturally high because the journals cover large areas of basic research with a “rapidly expanding but short lived literature that use many references per article” (pp. 498).

From another point of view, Moed and Van Leeuwen (1995) refer to the definition of citable documents, which exclude letters and editorials, and show that the impact factor is an inappropriate indicator, particularly for some journals. Favaloro (2008) further adds that impact factor tells little about the quality of a journal and Seglen (1994) indicates that the prestige or impact factor of a journal does not affect the strength level of its articles, however highly cited its articles are.

In contrast to the above mentioned studies, Franceschet (2010) considers two measures that are statistically correlated with each other: popularity and prestige[8] for the journals: the former counts the citations while the latter recursively weights the citations with the prestige of the citing journals. The situation is clear when both of them are high or both of them are low but there is some sort of controversy in the other two cases. He argues that popular non-prestigious journals are not necessarily highly cited and non-popular prestigious journals are not necessarily poorly cited. In another study, Laband and Piette (1994) show that sub-disciplines and their application matter to determine the number of citations regardless of research impact. This means that in some sub-disciplines, there is a naturally higher rate of citations.

Not only citation count is an important factor in scientific community, the number of published papers and scientists' productivity are other factors by which scientists are assessed. In terms of scientific productivity, it is measured by the number of articles. In general, governments and public authorities seek to maximize the benefit of public spending on research.

[8] Prestige is mainly related to journal ranking

Chapter 2 DATA SET AND RESEARCH METHODOLOGY

2.1 Introduction

In light of the literature reviewed in chapter 1, this chapter aims to shape the research method and address relevant methodology questions. As discussed in the previous chapter, paper production and research impact of academics have been discussed extensively and many determinants of this production are currently known as potential motives for scientific papers. Among others, age, gender, private and public funding, networking, field and context, university prestige, individual characteristics of researchers, and institutional setting are the most important factors.

In chapter 3, our general research question is as follows: do the mentioned factors (number of articles, amount of funding, size of research team, gender, and research filed) significantly affect citation counts? This comprehensive consideration of different factors in one study allows for testing the co-existing effect of different determinants, which provides an original contribution to the literature on citations.

In chapter 4, we tests whether collaboration with top-funded scientists shifts up scientist's number of articles. Collaboration with top-funded scientists can be an opportunity for accumulating valuable experience and tacit knowledge, resulting in higher and better scientific production. This objective can be reached by conducting a two-stage econometrics model that control the endogeneity issue.

We focus on the effect of holding a 'research chair' as a possible determinant of scientific publication in chapter 5. To find out the effect of holding a chair, we use a 'matching technique' and compare two 'chair-holder' and 'non-chair-holder' scientists who have similar funding, the same gender, and work in the same research field as each other.

The rest of this chapter is as following: section 2 presents the research questions and their linkages with the current literature. Section 3 presents the research objectives and discusses how these objectives will be achieved. Sections 4 and 5 review the data set and the econometric analysis respectively. Finally, section 6 reviews the contribution of this research to the advancement of knowledge.

2.2 Research questions

The first three questions belong to the research in chapter 3. Following the literature review in previous chapter, there may be numerous factors that identify the number of citations. We test whether the number of articles of a researcher in a given year or the journals' impact factor can influence his/her number of citations. Merton (1968) argues that scientists who publish more articles may be more visible to other members of scientific community and hence receive more citations. However, the visibility of researchers may initiate from other factors. Ale Ebrahim et al. (2014) highlights the role of publication marketing tools and strategies while Fowler and Aksnes (2007) refer to the self-citation, both of which improves the visibility of authors' prior works. We propose our first question:

Question 1: *Can higher visibility to scientific community, measured by the number of articles and the journal impact factors, positively affect the number of citations received in the future?*

Some evidence also mention that 'research collaboration' may affect 'research impact'. Godin and Gingras (2000) show positive correlation between citations and its collaborative nature. Scientific collaboration depends also on academic fields. Abramo et al. (2009b) show that in interdisciplinary studies there are more collaborations than strictly disciplinary research, and also that collaboration with foreign organizations are more common in basic fields than in applied ones. However, research collaboration is hard to be uniquely defined: First, it is a social relationship and it is not possible to identify the beginning and the end; second, determinants of scientific collaboration depend on the size of research team, institutional settings, domains, countries, and time-specific characteristics. Our second question relates to research team size and goes as follows:

Question 2: *Does the number of individuals in the list of authors significantly affect the future number of citations received by each scientist?*

There are some evidence in the literature about the positive effect of funding on scientific productivity (Arundel and Geuna, 2004; Harman, 2000; Pavitt, 2000, 2001). To formulate a question on the effect of funding, it can be inferred that funding is a necessary condition for scientific research production but that it is not a sufficient condition for generating high quality research. Research funding acts as a tool for researchers, and should be combined with researchers' skills and expertise to generate high quality results. From another perspective, it is possible to argue that funding is generally enough to contribute to the advancement of knowledge, but that it does not guarantee that this knowledge will be of high impact. Moreover, in laboratory-based fields, the role of funding may be more important in being able to perform research at all. The source of funding may also be important in affecting the scientific impact of research (Berman, 1990; Gulbrandsen and Smeby, 2005), as it shapes the project's goal and application. Hence our third question reads as follows:

Question 3: *Do the research funding from public, private for profit, and private not for profit sources have a positive effect on research scientific impact?*

In addition to the effect of funding, which has been discussed above, scientific collaboration and academic networking can also positively affect scientific productivity (Godin and Gingras, 2000). Melin (1996) also referred to the positive effect of co-authorship on scientific productivity at the local, national, and international levels. Combining the known effect of networking and scientific collaboration on the one hand and scientist funding capacity to generate knowledge on the other hand, it is interesting to investigate the effect of collaborators' funding on scientific productivity. To the best of our knowledge, there is not a comprehensive study on the effect of collaborators' funding on scientific productivity. Question 4 belongs to the research conducted in chapter 4. The chapter 4 tests whether collaborating with top-funded scientists has an effect on the scientific productivity measured by the number of articles. Hence our fourth question reads as following:

Question 4: *Does the collaboration with top-funded scientists positively affect scientific productivity?*

Being a top-funded scientist can be considered as a sort of prestige. However, prestige has broader and different meanings. The effect of prestige on scientific productivity has been already discussed in the literature (Frey and Rost, 2010; Long et al., 1979; West et al., 1998; Zhou et al., 2012). One measure of prestige can be honors and awards such as research chair. Cantu et al. (2009) showed the research chair program would be a good strategy for implementing knowledge-based development. In chapter 5, we investigate the questions 5 and 6, which are about the effect of holding a research chair on the number of scientific publication. Different types of research chairs are examples of awards. In Canada, there are three types of research chair: (1) research chairs which are awarded by industry and generally referred to as industrial chairs; (2) research chairs which are awarded by Canadian federal funding agencies such as the Natural Science and Engineering Research Council (NSERC) and the Canadian Institutes of Health Research (CIHR); and (3) the ‘Canada research chairs’, whose holders are assumed to already achieve research excellence in one of the main fields of research: engineering and the natural sciences, health sciences, humanities, and social sciences. Considering holding a chair as a measure of prestige, we aim to elucidate the effect of being a ‘chair-holder’ on scientific productivity. Our fifth question therefore reads as:

Question 5: *Does holding a chair increase a scientist’s productivity measured in terms of number of publications?*

Question 5 tests the productivity of chair-holders compared to other scientists and as such does not seek to prove causality. Considering the fact that chair-holders are the well-funded scientists too, this question cannot detach the funding effect of the chair from its other effects (mainly from prestige and networking effect). In other words, despite the evidence in literature about the benefits and goals of the research chair programs other than funding, question 5 is not able to disentangle effects.

It is possible to look at the research chair as a result of scientists’ characteristics and achievements (for instance age, number of articles, and number of citations). A chair-holder may experience an effect specific to holding chair on his/her scientific production in addition to the contribution of his/her characteristics and past achievements. To disentangle the exclusive effect of chair from the effect of scientists’ characteristics, we propose our sixth question:

Question 6: *Keeping a scientist’s main characteristics (age, research field, and amount of grants) constant, does holding a research chair have a significant positive effect on his/her scientific production?*

2.3 Research objectives

Having conducted this research, we may be able to meet the following objectives, starting from ‘building up a data set on funding and publication information of the Quebec researchers’ to investigating the above questions about the determinants of scientific productivity and research impact:

1- Build a comprehensive database on publication and funding information of Quebec scientists: this needs an appropriate mix of funding and publication information.
2- Measure the effect of visibility to the scientific community (measure by the number of past journal articles) and the journal impact factor, on the number of citations as a measure of research impact (**question 1**): this objective seeks for the effect of visibility on research impact.
3- Identify the effect of research team size (as a proxy of research collaboration) on the number of citations (**question 2**).
4- Explore the effect of the amount of funding on the number of citations (**question 3**)
5- Examine whether collaborating with a top-funded scientist may result in a higher number of publications (**question 4**)
6- Measure the correlation between 'holding a research chair' and scientific production, and investigate the causality enquireis (**questions 5 and 6**)

2.4 Data set

The data set used integrates information about researchers funding and publication in the province of Quebec. We have access to Thomson Reuters Web of Science database on scientific articles (2000-2012), which includes information about date of publication, journal name, authors, affiliations, and number of citation each article receives. Funding information of scientists comes from the Quebec University Research Information System (*Système d'information sur la recherche universitaire* or SIRU) of the Ministry of Education and Research (*Ministère de l'Éducation, de l'Enseignement supérieur et de la Recherche*). This database reports funding information including grants and contracts of all Quebec academics, on a yearly basis during the period 1985-2012. Age and gender of scientists was obtained from the ministry internal database.

We use variables which count the yearly number of articles [ln(*nbArticle*)], the yearly average number of authors in the papers of scientist [ln(*nbAuthor*)], and the five-year average of journal impact factor in which scientists publish [ln(*Impactfactor*)]. We also add dummy variables indicating the affiliated university of scientists. The next set of variables are the aggregate amount of money raised by each scientist. We separate the sources of research funding into public sector, private sector, or non-for-profit organization (NFP) with social and political mission. From another standpoint, research funding can have two purposes: it can be directly used for operation cost (O) such as research cost and researchers' salary (mainly student stipends and technical staff salaries) or it indirectly help research team buying instruments or larger infrastructure (I). In this framework, it is possible to have six variables for research funding for each researcher[9] [(1) ln(*PublicfundingO*); (2) ln(*PrivatefundingO*); (3) ln(*NFPfundingO*); (4) ln(*PublicfundingI*); (5) ln(*PrivatefundingI*); and (6) ln(*NFPfundingI*)][10]. In our dataset, we have used three-year average of funding to test the effect of past funding on current scientific productivity. Such three-year average takes all possible effect of past years into account. With the same rationale, we used the five-year average of journal impact factor.

The number of citations, varies from one discipline to another for two main reasons: (1) the number of papers and the amount of knowledge production is discipline dependent, and (2) citations may have different meaning in different disciplines. Two field-classifications are used in this paper. The first classification is used to categorize researchers into 9 broad domains, based on the U.S. classification of instructional programs[11]. The second classification is based on the field classification used by the U.S. National Science Foundation (||NSF). For instance, the citation patterns of papers in Economics is different from those of Political Science, despite the fact that both of them are subfields of Social Sciences. To control for such disciplinary-specific behaviour, we therefore calculate the relative number of citations [ln(*nbCitation*)], measured by the number of citations received so far by an article[12] divided by the average citation rate of the papers published in the same year in the same NSF speciality. This measure is then transformed by taking

---

[9] In the cases that the funding belongs to more than one researcher, the total amount of funding is divided by the number of researchers to make number more informative and reliable about its effect on scientific publication.

[10] In practice, we do not use *PrivatefundingI* and *NFPfundingI* in our analysis because of there too rare occurrence in our database. Funding for instruments or large infrastructure has a long-term effect that we are not able to capture with a maximum number of 12 years in the sample including the necessary lag structure. As will be explained in the methodology section ln(*PublicfundingI*) is used as an instrument for ln(*PublicfundingO*).

[11] The domains are Medical Science, Health Science, Business and Management, Social Science, Education, Humanities, Non-Health Professional, and Engineering, Science.

[12] More recent articles will have cumulated citations over a smaller number of years. To account for this fact, we add year dummy variables to the regression analysis.

its natural logarithm to normalize the variables and satisfy the necessary conditions for regression equations in this paper[13].

The pertinent variables to investigate the effect of collaboration with top-funded scientists identifies whether amongst the collaborator of an individual, there are co-authors who are in the top 10% and top 5% most funded individuals by field, distinguishing total funding, public sector funding, and private sector funding (it is also a kind of big-name or prestige effect). To generate the variables measuring collaboration with top-funded scientists, a set of dummy variables has been generated to identify whether a scientist has collaborated with a top funded scientist: [*ColT90*], and [*ColT95*] are the dummy variables that are equal to 1 if any of the coauthors in that year is amongst the top funded scientists (top 10%, and top 5% of total funding respectively). The funding amount used to generate these dummies is the sum of operational funding and funding for purchasing instruments. The variables of [*ColPub90*] and [*ColPub95*] are similarly built but only for public funding of coauthors and the variables of [*ColPriv90*] and [*ColPriv95*] have the same method of generation but for private funding of coauthors.

The titles of research project are being used to generate dummy variables identifying whether a scientist holds a research chair; the title field clearly states: "chair in…". As mentioned above, three types of chair are available in Canada: (1) industrial research chairs [*dIndChair*]; (2) research chairs awarded by Canadian federal granting councils [*dGCChair*]; and (3) Canada research chairs [*dCRC*]. In addition, we created a dummy variable [*dIndGCChair*] indicating whether the scientist is an industrial chair or a chair assigned by Canadian federal granting councils (the combination of *dIndChair* and *dGCChair*). Finally, the dummy variable [*dChair*] is equal to 1 for scientists with any type of chair (the combination of *dIndChair*, *dGCChair, and dCRC*). Building such dataset will satisfy the **objective 1**.

[13] There are four main assumptions in regression analysis regarding the data: (1) the variables should have normal distributions; (2) the linear relationship should exist between dependent and independent variables; (3) the variables should be reported without error; and (4) the variance of errors is the same across all levels of the IV (homoscedasticity). With the same justification regarding regression assumptions, the following variables in this paper are also transformed by logarithm function: funding, number of articles, number of authors and journal impact factor. For the variables that include a 0 value, we take the natural logarithm of 1 + the value of the variable.

2.5 Model and analysis

To investigate our research questions, we use regression analysis to estimate the number of publications or the number of citations by the variable listed in our research questions and a set of control variables discussed in the literature review chapter. The amount of financial support from different sources, the size of research team, the impact factor of journal, the age and the gender of scientists, university dummy variables, and field dummy variables are amongst the important variables. All of the questions can be validated by a panel regression model[14] that allows time variation as well as individual variation.

The goal of the sixth research question is to find the effect of holding a chair while chair and non-chair scientists are similar to each other in terms of some specific characteristics. In other words, we want to investigate what happens if two researchers with similar age, research fields, and funding amount are different from each other in terms of holding a research chair. Such comparison can be made by building the database of 'twin[15]' scientists with common characteristics and then comparing the effect of holding a research chair on scientific productivity in a panel regression model. This method is called matching technique.

The important concern about regression equations is the causality and bilateral correlation between dependent and independent variable. In a simple regression, to identify the determinants of scientific publication, the right hand side variable may be determined by the left hand side variable and this phenomenon is called endogeneity, which contradicts the assumption of a simple regression. In econometrics, a right-hand side variable is said to be endogenous if a correlation exists between that variable and the error term.

Endogeneity may occur for several reasons: measurement error, auto-regression with auto-correlated errors, simultaneity, omitted variables, and sample selection bias. In this study, some of the mentioned causes for endogeneity are potentially valid. First, financial support may be award based on the scientific production (simultaneity). Second, it is not possible to say all effective factors have been counted. Therefore, the existence of omitted variables is plausible. Third,

[14] Question 4 is about the prediction of a theoretical model, which should be validated by an empirical regression model.

[15] One of them should be chair and another one should be non-chair

scientific productivity cannot be fully reflected in the number of papers and patents and as a result, measurement error exists.

The 2SLS method of regression separates correlated part of right hand side variable and just keeps the independent part. In the first step, the endogenous independent variable in the regression equation is regressed on all other exogenous variables and instrument(s). An instrument is a variable that is not itself part of the explanatory equation and is correlated with the endogenous explanatory variables, conditional on the other covariates. In addition, the instrument cannot be correlated with the error term in the explanatory equation (conditional on the other covariates). This means that the instrument should not suffer from the same endogeneity problem as the original predicting variable. The predicted/estimated values are saved for the next stage. In the second stage, the normal regression will be done but the exogenous variables are substituted by the predicted values from the first stage. We can run a regression with 2SLS method to avoid such endogeneity.

A few variables are potential instruments for the amount of public funding. The number of scientists in a university [*nbScientistUni*] can explain the allocation of money amongst scientists. We anticipate that a university with a higher number of scientists may be able to benefit from cost sharing of research expenditures, hence reducing the need for larger amounts of individual funding. There is an established tradition in the literature to compare past research productivity of scientists to allocate funds (Ho et al., 2006; Liefner, 2003).

The rank of previous funding can be another choice to predicting the future amount of funding. There is an echo effect for the amount of funding, which means that highly funded scientists are better able to get new sources of research money. The logic behind this argument is that decent research funds have effective networking capacity (Winter et al., 2006) to create different opportunities to get funds in a country such as Canada (Salazar and Holbrook, 2007). It should also be noted that the rank of funding is an ordinal variable and as such does not have information about amounts (the amount of research fund may indicate the capacity of knowledge production while the rank of funding does not provide such information). Therefore, it is an informative variable to estimate the amount of funding but it is not the funding itself. The rank of a scientist in the field in terms of three-year average of funding for the purpose of operational costs and direct expenditure of research [*PubORank*] is the second instrument to predict the amount of funding.

The third instrument is the total funding of each research cluster in each university [*totPublicfundingOcluster*]. Having known the aggregate amount of funding, the rank of scientist in terms of funding, and the number of university, it is possible to make an estimation for funding of each scientist. In the first stage, the amount of public funding [*ln(PublicfundingO)*] is estimated by the instruments and the variables of the second stage regression. To avoid simultaneity problems, public funding is not contemporaneous to the instruments; hence one-year lags of the instruments are used in the first-stage regression. The equations used for regressions in chapter 3, 4, and 5 are in the following:

$$\underset{\text{Chapter 3}}{\ln\left(nbCitation_{it}\right)} = f\begin{pmatrix} \ln\left(PublicFundingO_{it}\right), \ln\left(PrivateFundingO_{it}\right), \ln\left(NFPFundingO_{it}\right), \\ \ln\left(nbArticle_{it}\right), \ln\left(ImpactFactor_{it}\right), \ln\left(nbArticle_{it}\right) \times \ln\left(ImpactFactor_{it}\right), \\ \ln\left(nbAuthor_{it}\right), dFemale_{i}, D_{Field}, D_{University}, D_{Year} \end{pmatrix}$$

$$\underset{\text{Chapter 4}}{\ln\left(nbArticle_{it}\right)} = f\begin{pmatrix} \ln\left(PublicFundingO_{it}\right), \ln\left(PrivateFundingO_{it}\right), \ln\left(NFPFundingO_{it}\right), \\ \ln\left(nbAuthor_{it}\right), ColTX_{it}, dFemale_{i}, Age_{it}, Age_{it}^{2}, D_{Field}, D_{University}, D_{Year} \end{pmatrix}$$

$$\underset{\text{Chapter 5}}{\ln\left(nbArticle_{it}\right)} = f\begin{pmatrix} \ln\left(PublicFundingO_{it}\right), \ln\left(PrivateFundingO_{it}\right), \\ \ln\left(NFPFundingO_{it}\right), \ln\left(nbAuthor_{it}\right), \\ (dIndChair \mid GCChair \mid dCRC \mid dIndGCChair \mid dChair)_{it}, \\ dFemale_{i}, Age_{it}, Age_{it}^{2}, D_{Field}, D_{University}, D_{Year} \end{pmatrix}$$

According to the chair characteristics in chapter 5, the networking and prestige effect of 'holding a research chair' may be mixed with the effect of funding. To address this issue, we use a matching technique, which is based on comparing two chair and non-chair scientists when they have some common characteristics. In general, matching is a statistical technique which is used to identify the effect of a treatment by comparing the treated and the non-treated individuals with similar observable characteristics. By matching treated individuals to similar non-treated units, it is possible to compare the effect of treatment, which is holding a research chair in this thesis.

Following the methodology employed by Bérubé and Mohnen (2009), it is possible to find pairs of chair-holders and non-chair-holders by using the psmatch2 command in Stata and then to remove the unmatched records. The selection is made by generating propensity scores and choosing the pairs of scientists with the closest scores to each other. The new data set consists of 'twin' scientists who are similar to each other in terms of funding, gender, and research fields. Controlling with funding, gender, and research field, and keeping only the matched scientists in the regressions, we are able to disentangle the prestige effect from the funding effect of the chair and hence, 'holding a research chair' becomes a more informative signal for the prestige of scientists. In this case, the effect of 'holding a chair' on scientific productivity does not include funding effect or it is not related to the field or gender of the scientist. One of the important check-point of the matching technique is validating the quality of matching. This implies that there should be no difference between the averages of the selection criteria (gender, funding, and research fields) when the comparison is made between chair holders and non-chair holders among the matched pairs. There can however be a difference when the comparison is made between the original database and the matched database.

2.6 Contribution to the advancement of science

Investigating the mentioned research questions brings some novel advancements in the analysis of scientists' productivity:

1- Mix of bibliometric and funding data in an econometric study: there are limited number of journal papers that investigate the effect of funding while they control other bibliometric indices. Such comprehensive empirical analysis improves the reliability, robustness, and significance of research results. We have access to Thomson Reuters Web of Science database on scientific articles (2000-2012), which includes information about date of publication, journal name, authors, affiliations, and number of citation each article receives. Funding information of scientists comes from the Quebec University Research Information System (*Système d'information sur la recherche universitaire* or SIRU) of the Ministry of Education and Research (MELS). This database reports funding information including research grants and contracts of all Quebec academics, on a yearly basis during the period 1985-2012. Age and gender of scientists was obtained from the MELS internal database. Description and summary of data are available in appendices.

2- Developing a theoretical model to predict the effect of collaboration with top-funded scientist and test the prediction with real data: this is not only an empirical study but also it is based on a theory model when it investigate the effect of collaboration with top-funded scientists. Inspired by Jensen and Meckling (1976) and by Durnev and Kim (2005), who developed concepts to model the incentives of firms' controller, we develop a theoretical model explaining the cost-benefit analysis of a researcher to publish papers.

3- Use of matching techniques to find out the effect of holding a research chair on scientific productivity: to show whether holding a research chair as an external support is important and significant in promoting scientific publication, we can run the two-stage panel regression on the entire data set. This determines whether 'holding a research chair' is a significant right hand side variable, either as a real cause or a channel for other variables/causes. However, the networking and prestige effect of 'holding a research chair' may be mixed with the effect of funding. To disentangle these two chair attributes, we use a matching technique and compare two 'chaired' and 'non-chaired' scientists who have quite similar amount funding, the same gender, and work in the same research field as each other. Following the methodology employed by Bérubé and Mohnen (2009), the selection is made by generating propensity scores and choosing the pairs of scientists with the closest scores to each other. The new data set consists of 'twin' scientists who are similar to each other in terms of funding, gender, and research fields. This technique improve the robustness and reliability of the results because it disentangles the prestige effect from the funding effect of the chair and hence, 'holding a research chair' becomes a better and more informative signal for the prestige of scientists.

4- A comprehensive investigation on determinants of research impact: to the best of our knowledge, there is not a study that identifies the determinant of scientific productivity in one model for the province of Quebec. This thesis provides a relatively comprehensive study on this topic.

Chapter 3 **ARTICLE 1 : WHAT DETERMINES RESEARCHERS' SCIENTIFIC IMPACT? A CASE STUDY OF QUEBEC RESEARCHERS**

Keywords: Bibliometrics, Science Policy, Citations, Scientific Publications, Scientific Impact

3.1 Abstract

Using a data set integrating information about researchers' funding and publication in the province of Quebec (Canada), this paper intends to identify the main determinants of citation counts as one measure of research impact. Using two-stage least square regressions to control for endogeneity, the results confirm the significant and positive relationship between the number of articles and citation counts. Our results also show that scientists with more articles in higher impact factor journals generally receive more citations and so do scientists who publish with a larger team of authors. Hence the greater visibility provided by a more prolific scientific production, better journals, and more co-authors, all contribute to increasing the perceived impact of articles. The paper also shows that male and female receive the same number of citations, all else being equal. In most of domains, the amount of funding does not have a significant effect on the citation counts. These results suggest that the most important determinants of researchers' citations are the journals in which they publish, as well the collaborative nature of their research.

3.2 Introduction

Research impact and its determinants are an important topic in science policy, as governments and public authorities seek to maximize the benefit of public spending on knowledge production and science advancement. Over the last decades, a vast body of literature has used citations to assess the scientific impact of scholarly research (Adam, 2002; Brown and Gardner, 1985; Kostoff, 2002; Narin, 1976). Although the classical interpretation of citations is that citations are building blocks in the construction of knowledge (Moed, 2005), many other factors have been shown to influence citation counts.

Assuming that citations can be a proxy for research impact (Kostoff, 1998; Moed, 2006; Phelan, 1999), it becomes important from a science policy perspective, to better understand the various factors that affect researchers' citation counts. The literature has highlighted several of the factors.

The size of research teams as an indicator of collaboration can influence research impact (Johnes, 1988; Melin, 1996). Similarly, previous findings shows that citation numbers depends on the domain, the prestige of the journal, and the social network of authors (Bornmann et al., 2008). Gender is also highlighted as a factor affecting the number of citations (Aksnes et al., 2011). The amount of research funding can be another determinant as it provides new opportunities for research and new sets of scientific findings (Harman, 2000; Pavitt, 2000, 2001), which could probably become an academic heritage.

We investigate the determinants of citations by analysing information about researchers funding and publication in the province of Quebec[16]. The first contribution of this paper is to analyze whether the various factors found in the literature are also positively associated, in our dataset, with changes in citation counts. Because those factors may be related with each other, the second contribution consists in collectively considering all of the mentioned factors in one model to test if their specific effects remain statistically significant. Here is our main research question: do the mentioned factors (number of articles, amount of funding, size of research team, gender, and research filed) significantly affect citation counts? This comprehensive consideration of different factors in one study allows for the testing of co-existing effect of different determinants and, as such, consists of an original contribution to the literature on citations.

The remainder of the paper goes as follows: Section 2 presents the conceptual framework and literature review; Section 3 explains the data set used in the paper and describes the research methodology; Section 4 analyses the regression results; and finally, Section 5 discusses the results and concludes.

3.3 Literature Review and Conceptual framework

Moed (2005) argues that citation is formal and based on pre-defined evaluation procedures, open without any restriction, and enlightening rather than formulaic. A few authors (for instance Cole and Cole (1971); Bornmann et al. (2008); Norris and Oppenheim (2003)) more or less verify the

[16] The dataset is presented in section 3.

correlation between the number of citations and research impact. However, important considerations in using citation counts must be highlighted.

A natural coherence between the number of citations and scientific impact generally implies that scientists select their references on the basis of the "impact" of the papers they cite, but this is not always the case. Authors sometimes cite papers to review the opposite view in the literature or to provide a general literature review (Amsterdamska and Leydesdorff, 1989; Seglen, 1997a). In another study, Moed et al. (1985) note that citations refer to impact on the scientific community and it does not completely reflect research "impact". The authors argue that any publication should thus have a minimum "quality" to impact other research but other factors like visibility of journals and the extent to which researchers provide a public service are two other important determinants for citing a particular paper that do not necessarily have a strong correlation with research impact. Later, Moed (2009) argues that other citation-based indicators (e.g. journal impact factors, Hirsch indices, and normalized indicators of citation impact) can be substitute indicators for measuring the scientific impact of research.

Along these lines, Kostoff (1998) investigates the theory of citation and suggests that every citation results from the combination of two main reasons: the real component of intellectual heritage and random components of self-interest. Although there is a random component, the author argues that the random effect disappears in the aggregation of citation counts and therefore the number of citations is a good indicator of the "quality" of research. Phelan (1999) provides the same justification. In a most recent review study by Bornmann and Daniel (2008), they argue that in most of citation-related studies, it is concluded that citing behavior is not only because of referring to intellectual and cognitive influences of other scientists and that there may be some non-scientific factors determining the decision to cite. However, the paper concludes that the different motivations of citers are "not so different or 'randomly given' to such an extent that the phenomenon of citation would lose its role as a reliable measure of impact" (pp. 45).

Assuming that citation counts are good proxies for the scientific impact of research, determinants of citation should be investigated. Some studies highlight a trade-off between "quality" and quantity of research (Broad, 1981; Butler, 2002; Hayes, 1983), which suggests that "high-quality" work takes time and, hence, cannot be produced at the same speed as less-quality work. In contrast, some studies argue that scientists who publish more articles may be more visible to other members of scientific community and such scientists may thus receive more citations—an indication of cumulative effects in science (Merton, 1968). There is some evidence in the literature indicating that a higher number of articles results in an increased number of citations in general (Feist, 1997; Hayati and Ebrahimy, 2009). Following this line of research, article counts can thus be considered as a proxy for the visibility of researchers (Aaltojärvi et al., 2008; Bar-Ilan et al., 2012).

However, the visibility of researchers does not solely depend on the number of articles they published and can stem from other factors. Ale Ebrahim et al. (2014) show that publication marketing tools and strategies significantly increase the article visibility and citation impact. Fowler and Aksnes (2007) indicate that self-citation improves the visibility of authors' prior works. It should be noted that the self-citation is not necessarily a negative point as it acts as a signal to readers about the author's prior work and background information (Sammarco, 2008).

In addition, publication in more prestigious journals (sometimes with higher impact factors) may provide a higher visibility for articles and hence gain more citations (Stegmann and Grohmann, 2001). Calderini and Franzoni (2004) and Seglen (1997b) argue that the impact factor of a journal may be one candidate to measure research impact or prestige. We test here whether the number of articles of a researcher in a given year or the journals' impact factor can influence his/her number of citations. With this evidence in mind, we propose our first hypothesis:

Hypothesis 1: *A higher visibility, measured by the number of articles and the journal impact factors, can positively affect the number of citations received in the future.*

Scientific collaboration depends on academic fields. Abramo et al. (2009b) show that in interdisciplinary studies there are more collaborations than strictly disciplinary research, and also that collaboration with foreign organizations are more common in basic fields than in applied ones. The number of individuals in a research team can be a proxy for the extent of networking as it shows the ability of researchers in collective scientific actions. More specifically, the number of staff in a department and the number of co-authors of articles at the local, national, and international levels can significantly explain research productivity of scientists (Johnes, 1988; Melin, 1996). However, Katz and Martin (1997) argue that co-authorship is just a partial indicator of collaboration. They also mention that the concept of 'research collaboration' is hard to define: First, it is a social relationship and it is not possible to identify the beginning and the end; second, determinants of scientific collaboration depend on institutions, domains, countries, and time. Our second hypothesis relates to research team size and goes as follows:

Hypothesis 2: *The number of individuals in the list of authors significantly affects the future number of citations received by each scientist.*

Funding and the effect of financial resources on generating research outputs have been investigated by a great number of scholars over the years (Arundel and Geuna, 2004; Harman, 2000; Pavitt, 2000, 2001). In most analyses, raising funds has been found to have a positive effect on scientific production. In terms of funding from the private sector, Gulbrandsen and Smeby (2005) find that industry funding is correlated with "new and interesting research topics and is prerequisite to accomplishing expensive and interesting projects" (pp. 947). The authors also indicate that professors with industrial funding collaborate more with other researchers, both from academia and industry.

Public funding can also have numerous effects on knowledge production. Beise and Stahl (1999) suggest the following six types of contribution publicly funded research may have: source of new useful knowledge, new instrumentation and methodologies, skills developed by those involved in carrying out basic research, expansion of national and international networks, dealing with complex problems[17], and creation of spin-off companies. However, it is not possible to claim the existence of a uniform and consistent argument about the effect of public funding; Appleyard (1996) indicates that public grants may have different effects on scientific production in two different countries or two different organisations.

To formulate a hypothesis on the effect of funding, it can be inferred that funding is a necessary condition for scientific research production but that it is not a sufficient condition for generating high quality research. Research funding acts as a tool for researchers, and should be combined with researchers' skills and expertise to generate high quality results. From another perspective, it is possible to argue that funding is generally enough to contribute to the advancement of knowledge, but that it does not guarantee that this knowledge will be of high impact. . Moreover,

[17] The paper argues that solving complex problems provides great benefit for the firms and organizations facing such problems.

in laboratory-based fields, the role of funding may be more important in being able to perform research at all. The source of funding may also be important in affecting the scientific impact of research, as it shapes the project's goal and application. Hence our third hypothesis reads as follows:

Hypothesis 3: *Research funding from public, private for profit, and private not for profit sources may have a positive effect on research scientific impact.*

The regression analysis described below will test whether having more financial support from the public, private and not-for-profit (NFP) sectors for the purposes paying research functional cost expenditures may increase the number of citations.

3.4 Data and Methodology

The data set used in this article integrates information about researchers funding and publication in the province of Quebec. We have access to Thomson Reuters Web of Science database on scientific articles (2000-2012), which includes information about date of publication, journal name, authors, affiliations, and number of citation each article receives. Funding information of scientists comes from the Quebec University Research Information System (*Système d'information sur la recherche universitaire* or SIRU) of the Ministry of Education and Research (MELS). This database reports funding information including grants and contracts of all Quebec academics, on a yearly basis during the period 1985-2012. Age and gender of scientists was obtained from the MELS internal database. Description and summary of data are available in the appendices section.

Our dependent variable, the number of citations, varies from one discipline to another for two main reasons: (1) the number of papers and the amount of knowledge production is discipline dependent (Moed et al., 1985), and (2) citations may have different meaning in different disciplines (Peters and Van Raan, 1994). Two field classification are used in this paper. A first classification is used to categorize researchers into 9 broad domains, based on the U.S. classification of instructional programs[18]. The second classification is based on the field classification used by the U.S. National Science Foundation (NSF). The latter field and subfield classification is used to perform the normalization of citations. For instance, the citation patterns of papers in Economics is different from those of Political Science, despite the fact that both of them are subfields of Social Sciences. To control for such disciplinary-specific behaviour, we therefore calculate the relative number of citations [ln(*nbCitation*)], measured by the number of citations received so far[19] divided by the average citation rate of the papers published in the same year in the same NSF speciality. This measure is then transformed by taking its natural logarithm to normalize it and satisfy the necessary conditions for regression analysis in the paper[20].

We use variables which count the yearly number of articles [ln(*nbArticle*)], the yearly average number of authors in the papers of scientist [ln(*nbAuthor*)], and the five-year average of journal impact factor in which scientists publish [ln(*Impactfactor*)]. We also add dummy variables indicating the affiliated university of scientists. The variables [ln(*nbArticle*)] and [ln(*Impactfactor*)] are used for testing hypothesis 1 and the variable of [ln(*nbAuthor*)] is used to test hypothesis 2.

The next set of variables addresses specific funding information on the amount of money raised by each scientist. We separate the sources of research funding into public sector, private sector, or non-for-profit organization (NFP) with social and political mission. From another standpoint, research funding can have two purposes: it can be directly used for operation cost (O) such as research cost and researchers' salary (mainly student stipends and technical staff salaries) or it indirectly help research team buying instruments or larger infrastructure (I). In our framework, it is possible to have six variables for research funding for each researcher[21] [(1) ln(*PublicfundingO*);

---

[18] The domains are Medical Science, Health Science, Business and Management, Social Science, Education, Humanities, Non-Health Professional, and Engineering, Science.

[19] More recent articles will have cumulated citations over a smaller number of years. To account for this fact, we add year dummy variables to the regression analysis.

[20] There are four main assumptions in regression analysis regarding the data: (1) the variables should have normal distributions; (2) the linear relationship should exist between dependent and independent variables; (3) the variables should be reported without error; and (4) the variance of errors is the same across all levels of the IV (homoscedasticity). With the same justification, the following variables in this paper are also transformed by logarithm function: funding, number of articles, number of authors and journal impact factor. For the variables that include a 0 value, we take the natural logarithm of 1 + the value of the variable.

[21] In the cases that the funding belongs to more than one researcher, the total amount of funding is divided by the number of researchers to make number more informative and reliable about its effect on scientific publication.

(2) ln(*PrivatefundingO*); (3) ln(*NFPfundingO*); (4) ln(*PublicfundingI*); (5) ln(*PrivatefundingI*); and (6) ln(*NFPfundingI*)][22]. The mentioned funding variables are used to test hypothesis 3.

Gender is also available in dataset [*dFemale*]. A great number of scholars have examined the gender effect on the number of publications and citations. Some articles show that women's publications are slightly less cited than that of men (Aksnes et al., 2011; Sugimoto et al., 2013). Long (1990) explains that women's opportunities for collaboration are significantly less than those of men's because women are more likely to have greater family responsibility than men. There are also some studies showing that women generally publish fewer articles than men do (Hesli and Lee, 2011; Leahey, 2006b). This may result in less visibility and hence fewer citations for women. However, Long (1992) finds that there is no difference between men and women in terms of the number of citations or level of contribution in a team's work. All being considered, the effect of gender on the number of citations should be tested and controlled in our model.

There are two possible levels to investigate number of citations: the author level or the article level. As we measure the effect of funding on relative citations, we are obliged to consider the author as the observation unit, rather than the article, as it is impossible to attribute a specific amount of funding to a particular article (even with using information contained in the acknowledgements). All variables are therefore aggregated (summed) at the scientist-year level (the database base being built as a panel).

Our regression analysis aims to understand the citation count determinants. It is important to note that the two variables of [ln(*PublicfundingO*)] and [ln(*nbCitation*)] are determined by each other simultaneously, and hence a potential source of endogeneity, which biases the ordinary least square (OLS) regressions. The main reason for this potential endogeneity is that the scientists are assessed for public funding based on their CV and past effectiveness while at the same time, publication and research impact depends on the funding capability of researchers.

[22] In practice, we do not use *PrivatefundingI* and *NFPfundingI* in our analysis because of there too rare occurrence in our database. Funding for instruments or large infrastructure has a long-term effect that we are not able to capture with a maximum number of 12 years in the sample including the necessary lag structure. As will be explained in the methodology section ln(*PublicfundingI*) is used as an instrument for ln(*PublicfundingO*).

Using instrumental variables (IV) is a standard technique suggested in literature to deal with such an endogeneity issue. Instruments must be correlated with the endogenous variable, and should not be correlated with the error term in the main regression equation, which implies that the instruments should not suffer from the same endogeneity problem. If there is more than one instrument for the endogenous variable, it is necessary to perform a two-stage regression, in which the first stage estimates the endogenous variable (named here as instrumented variable) on a list of instrumental variables. Such estimation removes the error term of the first stage and keeps the estimated amount for the second stage.

In the first stage, the amount of public funding [ln(*PublicfundingO*)] is estimated by the following variables: (1) the number of articles is an important factor that measures the past productivity of scientists [ln(*nbArticleAvg3*)] as it is the main component of one's CV; (2) infrastructure related public funding is a proxy indicating how much a scientist is equipped to conduct research in the frontier of knowledge [ln(*PublicfundingO*)][23]; (3) age and its square, which generally measures the a scientist' research experience [*Age*, Age^2]. There is evidence in literature about the non-linear effect of age (Bernier et al., 1975; Diamond, 1986; Kyvik and Olsen, 2008). It should be also noted that in first-stage regression, public funding is not affected by the number of articles in the same year but from the previous years. Hence a one-year lag of the three-year average of the number of articles is used as an instrument in the first stage regression. Using the same rationale, a one-year lag is also applied for the effect of infrastructure funding. The significance of the coefficients in the first stage regression (reported in the appendices section) and related tests show that these instruments are appropriate choices (they are significant).

Having the estimated amount of public funding from the first stage to tackle the endogeneity issue, the second stage estimates the relative number of citations received up to 10 years following the publication year [ln(*nbCitation*)] on funding [ln(*PublicfundingO*), ln(*PrivatefundingO*), and ln(*NFPfundingO*)], the number of articles [ln(*nbArticle*)], the average number of authors per paper [ln(*nbAuthor*)], and the five-year average of the journal impact factor [ln(*Impactfactor*)]. It should

[23] Infrastructure funding is generally a measure of the capability of generating original knowledge but when it comes from private or non-for-profit (NFP) sources, it is not significant as instrument and it only plays a small role in the development of infrastructure.

be noted that the funding variables are measured in three-year averages to smooth out large variations in yearly funding. The effect of gender is also tested.

Moreover, the interaction between the journal impact factor and the number of articles [ln(*nbArticle*) * ln(*Impactfactor*)] is added to investigate whether there is a moderating effect of the impact factor on the number of articles. The interactive variables may have additional and complementary explanatory power. For this paper, although the number of articles and journal impact factor may have an individual effect on the number of citations, their interaction can have significant collective effect.

In addition, we control for university effects to account for any impact that our explanatory variables do not cover. For example, papers from McGill University and the University of Montreal (UdeM) receive more citations (figure 3.1[24]) than those of other universities in the province. We also add year dummy variables to account for year-specific characteristics of the research system as exemplified by the evolution of citations over time (figure 3.2). The significant time trend and differences between universities justify the existence of these dummy variables in the model. The possible reason behind yearly differences is that research fluctuates each year based on the economy and research policy and such fluctuation may affect the research impact. University dummy variables can have the same role as research setting and related motivations are partially university dependent. Considering the mentioned explanatory variables, the resulting model is given by:

$$1^{st}\ stage: \ln\left(PublicFundingO_{it}\right) = g\left(\ln\left(nbArticleAvg3_{it-1}\right), \ln\left(PublicFundingI_{it-1}\right), Age_{it}, Age_{it}^{2}\right)$$

$$2^{nd}\ stage: \ln\left(nbCitation_{it}\right) = f\begin{pmatrix} \ln\left(PublicFundingO_{it}\right), \ln\left(PrivateFundingO_{it}\right), \ln\left(NFPFundingO_{it}\right), \\ \ln\left(nbArticle_{it}\right), \ln\left(ImpactFactor_{it}\right), \ln\left(nbArticle_{it}\right)\times\ln\left(ImpactFactor_{it}\right), \\ \ln\left(nbAuthor_{it}\right), dFemale_{i}, D_{Field}, D_{University}, D_{Year} \end{pmatrix}$$

[24] The small universities are grouped according to their institutional similarities. The University of Quebec (except the one in Montreal) and Bishop University are in the same group. The second group includes “École de technologie supérieure” (ETS), “Université du Québec à Montréal” (UQAM), and “Institut national de la recherche scientifique” (INRS).

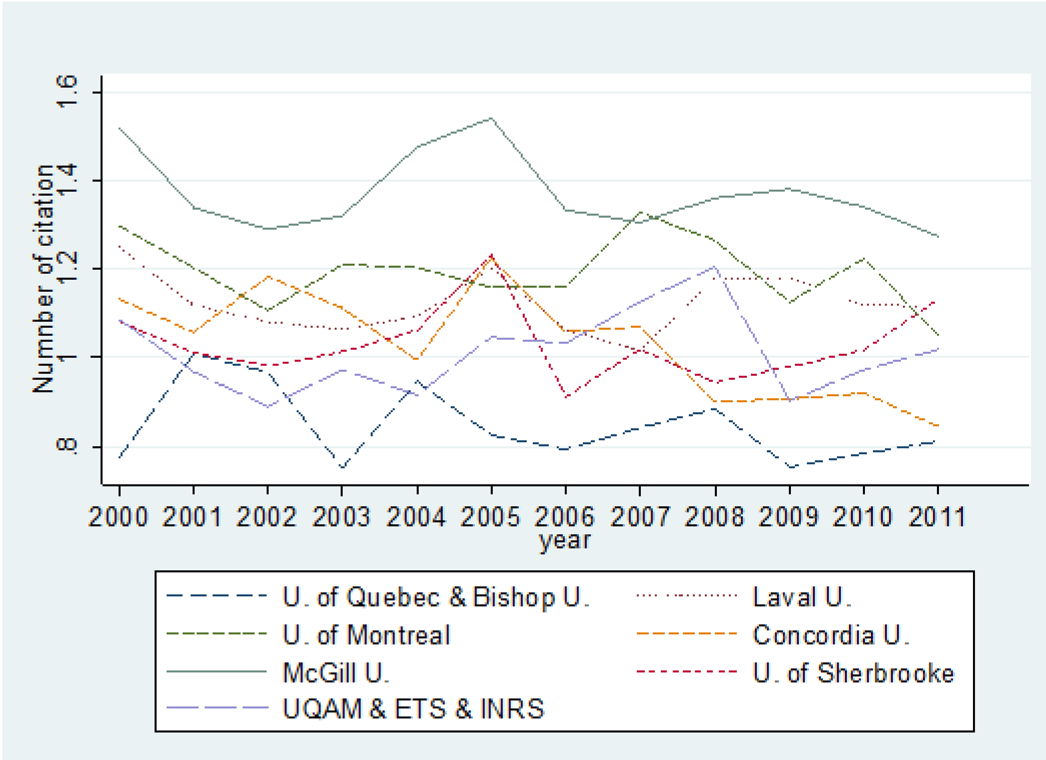

Figure 3.1: Discipline-normalized citation rates of papers from Quebec universities

Figure 3.2: Average of Discipline-normalized citation rates of Quebec papers, by year

For a panel data set, the 'xtivreg' Stata command can be used to estimate two-stage least square models (2SLS)[25]. The regression analysis is conducted for each domain separately. This results in a better understanding of determinants specific to different domains. Just focusing on the regression analysis for entire data set may generate some holistic arguments without any domain-specific interpretation. For instance, funding is not similar across domains, nor is the average number of authors per article. As the sample size per discipline precludes estimating the regressions for each discipline, the compromise used in this paper is to run regressions on groups of disciplines. Description and summary of variables are available in the appendices section.

[25] Because we are using relative citations and not direct counts of citations, Poisson or negative binomial regressions are not appropriate.

3.5 Regression analysis

Because a number of our independent variables are individual effect, which are constant over time (for instance gender or university affiliation), we therefore prefer to estimate random effect 2SLS regressions, which are reported in table 3.1 (first stage regression results are reported in the appendices section). However, to check for the robustness of regression results, the data will be treated for both cross section and panel data. In addition to the two-stage regression (*xtivreg* command in Stata), a simple ordinary least square (OLS) without endogeneity is also tested for both cross section and panel data (*reg* command and *xtreg* command in Stata). The cross section OLS (no consideration of time effect) and Panel OLS analysis are reported in the appendices section.

Table 3.1: The second stage regression result (citation count)

Dependent variable: ln(*nbCitation*)$_{it}$	**Domain**												
	A	**B**	**A+B**	**C**	**D**	**C+D**	**E**	**F**	**G**	**H**	**I**	**H+I**	**All**
dFemale$_i$	0.0095	0.0033	0.0073	0.0396	-0.0019	0.0055	-0.1010	-0.1614**	-0.0820	0.0109	-0.0245	-0.0140	-0.0063
	(0.0134)	(0.0231)	(0.0116)	(0.0331)	(0.0221)	(0.0188)	(0.1007)	(0.0722)	(0.0622)	(0.0298)	(0.0193)	(0.0161)	(0.0077)
ln(*nbArticle*)$_{it}$	0.0329***	0.0985***	0.0422***	0.0985**	0.0498**	0.0464**	0.0127	-0.0129	0.0886*	0.0845***	0.0604***	0.0679***	0.0570***
	(0.0107)	(0.0223)	(0.0097)	(0.0450)	(0.0234)	(0.0211)	(0.1274)	(0.1593)	(0.0520)	(0.0149)	(0.0119)	(0.0092)	(0.0059)
ln(*Impactfactor*)$_{it}$	0.2987***	0.2522***	0.2875***	0.3815***	0.2678***	0.3026***	0.2612*	0.4352**	0.0715	0.2462***	0.2301***	0.2370***	0.2458***
	(0.0207)	(0.0474)	(0.0189)	(0.0692)	(0.0367)	(0.0319)	(0.1493)	(0.1702)	(0.0891)	(0.0294)	(0.0242)	(0.0185)	(0.0105)
ln(*nbArticle*)$_{it}$*ln(*Impactfactor*)$_{it}$	0.1058***	0.0934**	0.1062***	-0.0184	0.0254	0.0046	0.0947	-0.1007	0.1765*	0.0330	0.0738***	0.0592***	0.0901***
	(0.0171)	(0.0431)	(0.0158)	(0.0784)	(0.0392)	(0.0345)	(0.1754)	(0.2154)	(0.0990)	(0.0264)	(0.0206)	(0.0160)	(0.0094)
ln(*nbAuthor*)$_{it}$	0.1945***	0.0585***	0.1785***	0.0755**	0.0959***	0.0894***	0.0612	0.0264	0.2066***	0.1599***	0.1296***	0.1389***	0.1357***
	(0.0090)	(0.0203)	(0.0083)	(0.0362)	(0.0168)	(0.0150)	(0.0844)	(0.0720)	(0.0435)	(0.0153)	(0.0089)	(0.0076)	(0.0046)
ln(*PublicfundingO*)$_{it}$	-0.0044	-0.0085	-0.0062	0.0084	0.0270**	.02068*	-0.0328	0.0462	0.0240	0.0109	0.0302***	0.0237***	-0.0012
	(0.0045)	(0.0133)	(0.0043)	(0.0193)	(0.0133)	(0.0113)	(0.0456)	(0.0647)	(0.0332)	(0.0128)	(0.0082)	(0.0069)	(0.0034)
ln(PrivatefundingO)$_{it}$	0.0028***	-0.0002	0.0025***	-0.0073	0.0012	-0.0015	0.0024	-0.0338	-0.0093	-0.0016	-0.0032**	-0.0031***	0.0001
	(0.0010)	(0.0028)	(0.0009)	(0.0050)	(0.0036)	(0.0029)	(0.0193)	(0.0222)	(0.0065)	(0.0021)	(0.0015)	(0.0012)	(0.0006)
ln(*NFPfundingO*)$_{it}$	-0.0014	0.0010	-0.0009	0.0057	0.0005	0.0025	0.0086	-0.0135	-0.0028	0.0011	-0.0006	0.0002	-0.0004
	(0.0011)	(0.0024)	(0.0010)	(0.0044)	(0.0025)	(0.0022)	(0.0134)	(0.0155)	(0.0082)	(0.0019)	(0.0014)	(0.0011)	(0.0006)
Constant	0.2381***	0.3000**	0.2526***	0.0673	0.0294	0.0764	0.8766***	-0.0276	-0.0610	-0.0174	-0.0405	-0.0334	0.2489***
	(0.0416)	(0.1207)	(0.0398)	(0.1619)	(0.1095)	(0.0929)	(0.3144)	(0.4933)	(0.2976)	(0.1223)	(0.0763)	(0.0649)	(0.0309)
Number of observations	9026	1761	10787	1071	2944	4015	253	369	547	4029	7261	11290	31563
Number of scientists	1270	301	1571	301	673	974	104	202	178	664	1100	1764	5387
$\chi 2$	3665***	578***	4188***	462***	811***	1188***	84***	134***	182***	941***	1722***	2695***	9448***
Average year activity	7.11	5.85	6.87	3.56	4.37	4.12	2.43	1.83	3.07	6.07	6.60	6.40	5.86
R^2 within groups	0.25	0.22	0.24	0.20	0.14	0.16	0.27	0.24	0.20	0.15	0.13	0.14	0.18
R^2 overall	0.32	0.27	0.31	0.31	0.24	0.26	0.29	0.28	0.26	0.23	0.23	0.23	0.27
R^2 between groups	0.47	0.24	0.43	0.39	0.35	0.35	0.19	0.25	0.24	0.28	0.38	0.35	0.34

*, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively – Year dummies and university dummies are significant. The definition of regression tags indicating the sample: A= Medical, B= Health Science, c= Business and Management, D= Social Science, E= Education, F= Humanities, G= Non-health professions, H= Engineering, I= Science.

The focus of this section is to analyse the two-stage least square regression (table 3.1) which addresses the problem of endogeneity. The regressions imply that, *ceteris paribus*, scientific publications of female scientists are cited in the same manner as men's publication. The variable of [*dFemale*] is not significant but there is a negative effect of being female only in the humanities (the small sample size and the fact that research output is badly measured by the Web of Science, we will not dwell on this result). This finding seems to be different from what Aksnes et al. (2011) and Larivière et al (2013) show about the underperformance of women.

The number of articles [ln(*nbArticle*)] has a significant positive effect except for education and the humanities. Hence, in general, a greater visibility of scientists, as determined by their number of articles, is associated with a greater citation rate.

Having added the interaction between journal impact factor [*ln(Impactfactor)*] and number of articles [ln(*nbArticle*)] to the regression, the results still show that not only there is a positive effect of articles count (due to the author's visibility), but also those papers in higher impact factor journals receive more citations (all domains except non-health professions). In addition, a higher number of articles in more higher impact factor journals (interactive variable) results in more citations than the same number of articles in a less prestigious journal (this effect is significant for medical, health science, non-health professions, and science). Consequently, it is possible to argue that hypothesis 1 becomes validated as the higher visibility, measured by the number of articles and the journal impact factors, can positively affect the number of citations.

Turning now to the size of research team, our results show that articles with more authors [ln(*nbAuthor*)] are more likely to be cited. It indicates that hypothesis 2 is validated. This finding is compatible with evidence in literature indicating positive effect of collaboration on research impact (Johnes, 1988; Melin, 1996). As a justification, in a research team with numerous researchers, tasks are done collectively and by different scientists. It probably provides some sort of knowledge spillover or tacit knowledge transfer, which improves capability of researchers in conducting high impact research.

Moreover, funding does not have a major effect on the number of citations. For example, [ln(*PublicfundingO*)] has a positive significant effect only for science and social science. The effect of [ln(*PrivatefundingO*)] is only significant for the medical (positive effect) and science (negative effect) domains. The coefficient of [ln(*NFPfundingO*)] is not significant at all. The empirical and theoretical evidence in literature supporting the positive effect of funding on productivity of scientists are known (Arundel and Geuna, 2004; Harman, 2000; Pavitt, 2000, 2001), but our results indicate that higher funding does not necessarily results in publications which are more cited. This paper does not contradict the positive effect of funding on scientific productivity but it indicates that higher funding is not a determinant of article citation counts. As a result, hypothesis 3 cannot be validated because there is no significant and comprehensive evidence for positive effect of funding from different sources (public, private, or NFP organization) on research impact.

3.6 Policy Implications

Assuming that the number of citations is a good proxy for research impact and, in turn, for a certain kind of *quality*, we propose some policy advice to address the issues discussed in the paper. First, it seems that collaborative works (measured by the size of research team) can influence the impact of the research, and policy makers should therefore encourage research of a more collaborative nature. The measure of such research collaboration is not only limited to the number of authors in each articles but can also measure the impact, extent, and durability of collaboration. However, the only variable we have on hand, which measure the research collaboration, is the number of individuals in the authors list.

Second, the significance level of the funding effect on research impact (which is not the same for all domains) does not necessarily imply that funding is ineffective for the knowledge production process/chain – this paper only investigates researchers' scientific impact and not their research productivity, which his considered as an input here. There is strong evidence in the literature about the significant effect of different funding types on scientific productivity (Manjarrés-Henríquez et al., 2008; Pavitt, 2001; Salter and Martin, 2001) to which our research contributes: public funding has a positive and significant effect for "Science" and "Social Science", while the private funding effect is positive and significant for "Science" and "Medical Science". These results point towards domain-specific policies and incentives because the effect of funding is domain-dependent, *ceteris paribus*. Hence a domain-specific policy can be an effective tool for improving the research impact in specific domains, without the need for general policy making, which may require holistic manipulation of science policy.

The third policy implication is to incentivise researchers to publish in journals with a high impact factor. Such journals have more visibility and their articles are widely read and used by the scientific community, more so than articles in other journals with lower impact factor. Moreover, the impact factor of a journal is a proxy for research impact because journals with higher impact factor have more submissions and editors are able to choose higher quality papers. Although a greater number of articles in the past contributes to improving the visibility of articles in the future (and hence their perceived research impact), the positive and significant effect of the interactive variable, which measures the modulating effect of the number of articles on the journal impact factor, on research impact implies that past articles in journals with a higher impact factor can reflect the intrinsic research impact of individuals. We can also argue that there is a learning experience from the past collaboration, especially if that collaboration led to a highly cited paper.

The last but not the least, our research did not find any gender bias in terms of research impact (except for the domain of Humanities). This contrasts with some evidence in literature that points towards the relative under-performance of women in terms of number of articles and research impact. However, it is not possible to make a policy conclusion in this regard because there may be a great number of reasons that may explain why women are less cited in one specific domain. First, they publish less and are thus less visible. Second, there may be conscious or unconscious discrimination. Third, women may be involved in more multidisciplinary research that is harder to publish. Fourth, women may spend more time involved in other duties at university. Although more investigations are needed, some incentive programs can be to encourage women to apply for more funding, and to go to more conferences (to be more visible). In addition, mentorship programs should be put in place where the gap is significant.

3.7 Conclusion

The paper investigates the determinants of citation counts as an indicator of research impact. To reach that aim, three propositions were set to be validated: one on the effect of funding, one on the effect of research team size, and another one on the effect of articles count and journal impact factor. The last two have been validated and the first one is rejected. In conclusion, this paper shows that the number of articles and the visibility of a researcher, the impact factor of the journal, the size of the research team, and the institutional setting of the university are the important determinants of citation counts. In addition, the regressions show that there is no significant effect of public funding and gender in most of the domains examined. Furthermore, it should be noted that for the domains of Education, Humanities, and Non-health professions, hypothesis 1 (about the effect of articles count and journal impact factor) is not validated. Moreover, hypothesis 2 (about the effect of research team size) for the domains of Education and Humanities is not validated.

In terms of validity of abovementioned interpretations, it should be noted that the study only covers Quebec scientists and some entries are missing in original dataset. In addition to using more comprehensive and complete data set, there are also some other suggestions for future works in this subject. First, it would be interesting to investigate the citations and group them to distinguish self-citations, citations due to high research impact, and citations for bringing evidence from literature. Each group of citations may have different sets of determinants. Second, it would also be of importance to investigate the time trend of citation whether it is possible to make some arguments about timing of citation accumulation of each scientist or each paper, based on different events and different factors.

Chapter 4 ARTICLE 2 : THE EFFECT OF COLLABORATION WITH TOP-FUNDED SCIENTISTS ON SCIENTISTS' PERFORMANCE

Keywords: Scientist Productivity, Publication Determinants, Research Funding, top-funded Scientists

4.1 Abstract

The theoretical model developed in this paper predicts that collaboration with top funded scientists positively affects the number of scientific publications. Having combined data on funding and publication of Quebec scientists, this paper empirically tests the theory predictions. The paper examines numerous definitions of top-funded scientists as those in the top 25%, top 10%, or top 5% in terms of total funding, funding from the public sector, and funding from the private sector. The results show that collaborating with such top-funded scientists has a positive effect on scientists' number of publication, which is compatible with our theory predictions.

4.2 Introduction

Scientists' academic productivity has been extensively discussed in and out of the literature and many of its determinants are currently known as potential motives for publishing papers in peer reviewed journals. Among others, age, gender, private and public funding, field and context are the most important factors. Funding definitely plays a major role in shaping scientific productivity and has been extensively investigated in the literature. A number of scholars show the positive effect of funding on academic productivity (Crespi and Geuna, 2008; Pavitt, 2000, 2001; Salter and Martin, 2001). Others find evidence on the effect of different types and sources of funding, some of which indicate that the effect of public funding and private funding are both positive (Berman, 1990; Gulbrandsen and Smeby, 2005). Other studies find that funding from the private sector has a detrimental effect on scientific productivity (Goldfarb, 2008; Kleinman and Vallas, 2001).

In addition to funding, scientific collaboration and academic networking can also positively affect scientific productivity. Using data on collaborative research conducted in Canada, Godin and Gingras (2000) indicated that not only is there a non-negative correlation between research impact and its collaborative nature, but that new research opportunities and scientific productivity are developed during research collaboration. Melin (1996) also referred to the positive effect of co-authorship on scientific productivity at the local, national, and international levels. Research collaboration is also a positive determinant of patenting in addition to scientific publication. Azoulay et al. (2007) provided some evidence that having a co-author who patented before “increases the likelihood of a patent application” (pp. 601) and also that scientists in universities that have good patenting records, generally patent more. In another study, Zhou (2003) also identified the importance of having a creative co-worker in leading to increased creativity for a researcher.

The networking effect can also be investigated from the standpoint of peripheral support in universities and research institutions. For instance, (Crane, 1965) highlighted the effect of university prestige on research productivity and Johnes (1988) indicated that the number of staff in a department and the number of co-authors of published articles can significantly explain the research productivity of UK economists in academia.

Combining the known effect of networking and scientific collaboration on the one hand and scientist funding capacity on the other hand, it is interesting to investigate the effect of collaborators’ funding on scientific productivity. Collaboration with top-funded scientists can be an opportunity for accumulating valuable experience and tacit knowledge, resulting in higher and better scientific production. This study thus tests whether collaboration with top-funded scientists shifts up scientist’s productivity. The rest of paper is organised as follows: A brief literature review and the theoretical framework are presented in section 2, the data and the econometric model are described in section 3, analysis of econometric results is the focus of section 4, and finally section 5 concludes.

4.3 Theoretical model and empirical framework

In this section, we develop a theoretical model to show how the collaboration with top-funded scientists results in a higher number of peer reviewed journal articles. A brief review of the empirical literature is also provided to show its compatibility with our theoretical model. Inspired by Jensen and Meckling (1976) and by Durnev and Kim (2005), who developed concepts to model the incentives of firms' controller, the following paragraphs model the incentives of a researcher to publish a paper. This model explains the cost-benefit analysis of a researcher to publish paper.

A number of assumptions are required prior to developing the model:

1- A scientific collaboration between two researchers affects the research results. The power and quality of a scientist's collaboration in terms of production of publishable knowledge is represented by a constant 'α' that is set between 0 and 1. In other words, 'α' is the probability of acceptance of a submitted paper (when focusing on the number of papers) or the level of its impact on future research in the field (when focusing on the impact of papers).

2- The maximum contribution of all possible external factors of publication is represented by $\overline{\pi}$. In other words, it is the 'maximum external potential of knowledge production' (consequences of the factors affecting scientific production such as research budget or research support from the university at organizational level).

3- The utility function of researchers is linear and risk neutral. A researcher compares the utility gained by publishing with the research cost and salaries paid to research assistants. We suppose that every unit of scientific outcomes, which is published, decreases the external potential of the knowledge production. In addition, such external potential should be summed up with personal willingness/ability of a scientist to publish a paper. This sum shows the aggregate possibility of scientific publications. In our utility function, this willingness is normalized to 1. The j^{th} unit of external and personal scientific potential is multiplied by α, the factor of collaboration effectiveness to calculate the gained utility from the j^{th} article. Then it is compared with the constant cost of publication (c) to determine whether it is worth publishing or not.

4- If the j^{th} article is being published, the following inequality should be valid, indicating that the benefit of a publication should be more than its cost. The left hand-side of the inequality represents the utility/benefit/quality of one unit of scientific outcome or knowledge when it becomes a paper. The right hand-side of the inequality represents the cost of publication.

$$\alpha\left(1+\pi(j)\right) > c$$
$$where\ \pi(j) = \overline{\pi} - j$$

5- Our last assumption states that if there is no support for a scientist $\left(\pi(j)=0\right)$, publishing is not possible ($c > \alpha$). It means that a scientist cannot publish only based on personal characteristics and without getting support from the academic community.

The next step consists in finding the optimum number of papers $\left(j^*\right)$. The utility gained from the j^{th} publication is decreasing in the number of articles. Therefore, j increases until the inequality becomes equal.

$$j^* = \frac{\alpha\left(1+\bar{\pi}\right)-c}{\alpha}$$

If j^* is divided by the 'maximum external potential of knowledge production' $\left(\bar{\pi}\right)$, we can calculate the percentage of external scientific potential which has been published as papers p^*. This p^* is a standardized level of publication regardless of the amount of external support and funding. It is a more appropriate tool for our analysis than j^* because j^* cannot be compared among the different types of external support that the scientists receive.

$$p^* = \frac{\alpha\left(1+\bar{\pi}\right)-c}{\alpha\bar{\pi}} = 1+\frac{\alpha-c}{\alpha\bar{\pi}} = 1+\frac{1}{\bar{\pi}}-\frac{c}{\alpha\bar{\pi}}$$

It is possible to find the sensibility of p^* to factors such as $\bar{\pi}$, c, and □ to extract the theoretical hypotheses that will be tested empirically further in the paper. The above model suggests the following statements:

1- Greater institutional and peripheral support results in more publications. This statement results directly from our theoretical model:

$$\frac{\partial p^*}{\partial \bar{\pi}} = -\frac{\frac{\alpha-c}{\alpha}}{\bar{\pi}^2} > 0$$

2- Researchers with higher costs of publication are less likely to publish:

$$\frac{\partial p^*}{\partial c} = -\frac{1}{\alpha\bar{\pi}} < 0$$

3- Higher 'power and quality of scientific collaboration' (higher α) improves scientific productivity. Higher α in our model refers to collaboration with top-funded scientists.

$$\frac{\partial p^*}{\partial \alpha} = -\frac{c}{\alpha^2 \overline{\pi}} > 0$$

This study then aims to empirically investigate the effect of collaboration with top-funded scientists on the number of scientific publications, which is related to the first and third predictions of our theoretical model. Acknowledging the proximity between funding and publishing, we consider that top-funded scientists are probably also star scientists.

There are numerous definitions of star scientists in the literature. For instance, Higgins et al. (2011) considered the recipients of a Nobel Prize as star scientists and showed that if they are affiliated to a biotechnology company they act as a signal of research impact to absorb resource and investment. The authors justify this by reasons such as Noble prize winners transfer tacit knowledge or bring valuable network to the company. In another study, Meyer (2006) proposed a different definition of star scientist (one who is active in both patenting and publication) and showed that inventor-authors' publications are over-proportionally high with a high number of citations, suggesting that the combination of publication and patenting activities is a good proxy for being a star scientist.

We propose to also consider top-funded scientists as stars. This would be reasonable enough because of the substantial effect of funding on scientific productivity from both qualitative and quantitative points of view. For instance, Beise and Stahl (1999) show the positive effect of publicly funded research on developing new products and processes. Pavitt (2001) also refers to the importance of public support for scientific infrastructure development and highlights its role in the effectiveness of public grants. In another study, Pavitt (2000) argues that infrastructure of expertise, equipment and networks is necessary for development and implementation of research, which thus is costly. In a review article of the effect of funding, Salter and Martin (2001) suggested the following six types of contribution of publicly funded research: new knowledge, advance instrumentation and methodologies, skills to conduct basic research, expansion of research networks, dealing with complex problems, and establishing spin-off companies.

The literature also provides evidence to the effect that research collaboration is an important determinant of scientific productivity. Scientific collaboration and being an active member of a scientific network positively affect scientific productivity (Azoulay et al., 2007; Chwe, 2000; Godin and Gingras, 2000; Hicks, 1995; Johnes, 1988; Melin, 1996). Scientific collaboration, however, depends on the academic field. Abramo et al. (2009b) showed that in interdisciplinary fields there are more collaborations than in mono-disciplinary fields and also that collaboration with foreign organizations more often occur in the basic fields than in the applied fields.

Scientific collaboration can also be analysed from a social network point of view. Chwe (2000) analysed social networks using a game theoretic model and identified three levels of hierarchy as the minimal sufficient network structure for efficient collaboration. These hierarchical stages are 'initial adopters', 'followers', and then 'late adopters'. The author highlighted that "a communication network helps coordination in exactly two ways: by informing each stage about earlier stages, and by creating common knowledge within each stage" (pp. 1). The implication of such structure in the scientific community would thus be a network in which scientists collaborate and produce new outputs, supposing that they know past information.

A few authors have considered the concept of 'research collaboration' with 'star scientist'. Azoulay et al. (2010) found that when a star scientist unexpectedly dies, his/her coauthors are faced with a decline in their quality-adjusted publication rates. The authors also noted that such a decline is due to the shortage of knowledge spill over between coauthors. In studying the impact of co-authors, Oettl (2009) proposed the concept of helpfulness in addition to productivity to classify a group of immunologist star scientists: high productivity and helpfulness (All-stars), high productivity but average helpfulness (Lone Wolves), average productivity but high helpfulness (Mavens), average productivity and helpfulness (Non-Stars). His results showed that after the death of star scientist, the output of the coauthors dropped by 16% in the case of All-star and by 14% in the case of Mavens. In the other two cases, the outputs remained constant. Such results showed that considering only productivity may be systematically wrong as the Lone Wolves are overvalued and the Mavens are undervalued.

Beaudry and Schiffauerova (2011) showed that in the Canadian biotechnology sector the presence of more star scientists in the research team has a positive effect on patent quality. Working on the same database, Schiffauerova and Beaudry (2011) noted that star scientists play the role of knowledge gatekeeping who acquire external knowledge and nurture the clusters because they have numerous numbers of connections outside and inside clusters, enabling them to play such a role. Regarding the contribution of star scientists in industry, additional evidence showed that the involvement of a star scientist in the firm positively affects how successful the firm is (Bagchi-Sen, 2007; Bercovitz and Feldman, 2011; Darby and Zucker, 1996; Rothaermel and Hess, 2007; Zucker and Darby, 1996, 1997, 2001; Zucker et al., 2002).

Both the effects of funding and collaboration on scientific productivity have been well investigated in literature. To the best of our knowledge, however, there is not a comprehensive study on the effect of collaborators' funding on scientific productivity. Based on the known correlation between funding and scientific productivity, we consider funding as a proxy of being star. In other words, being a top-funded scientist is an instrument showing the intrinsic skills and abilities of scientist. This research tests whether collaborating with top-funded scientists has an effect on the scientific production measured by the number of articles. Hence our hypothesis reads as follows:

Hypothesis:

Collaboration with top-funded scientists positively affects scientific productivity.

Other determinants of scientific productivity have been investigated in the literature, and they should therefore be considered in our empirical analysis as control variables. For instance, gender is known as a significant determinant of scientific productivity in the literature. Some researchers argue that differences in research productivity between men and women come from administrative positions of researchers and their marital status. Xie and Shauman (1998) indicated that "women scientists publish fewer papers than men because women are less likely than men to have the personal characteristics, structural positions, and facilitating resources that are conducive to publication" (pp. 863). Xie and Shauman (1998) nonetheless concluded that gender differences in research productivity has declined over time, while at the same time the population of female scientists has proportionally increased - this decline is also observed in Abramo et al. (2009a). Fox (2005) argued that the effect of gender is complex in a way that it is not possible to simply separate the effect of married and single researchers and he refers to the career of the spouse and to the family composition as two important factors of such complexity. For example, women with preschool children show higher productivity than women without children and women with school-age children. Nakhaie (2002) also investigated the gender effect in different durations of time and argued that Canadian female professors publish less than their male colleagues do, both in a lifetime period and during a shorter period, but that such effect is higher for the former.

Age is another independent determinant of scientific productivity examined in the literature. There is evidence that the effect of age varies between different disciplines (Kyvik, 1990). The authors indicated that the productivity level in the social sciences is independent of age. In the humanities, however, publishing activity declines during the age-period of 55-59, but it is followed by a new peak in the group of 60 years old and over. There is a different story in the medical sciences in which the productivity falls when researchers reach 55 years and older. In the natural sciences, productivity continuously decreases with aging because new scientific tools, methods, and equipment are continuously introduced and older researchers may have problems becoming familiar with them (Kyvik, 1990).

The inverted U-shaped effect of age (Quadratic or second-order effect of age) receives significant support in the literature. In a comprehensive study, Kyvik and Olsen (2008) tried to justify the hypotheses in the literature explaining the age effect on academic productivity and categorized them in six groups. All of the following hypotheses have been locally verified based on the data set used: (a) The utility maximizing hypothesis implies that academic staff conduct less research as they age because the expected utility of time spent on research diminishes; (b) The seniority burden hypothesis refers to the increasing administrative load as a career advances, which decrease the focus on scientific matter; (c) The cumulative disadvantage hypothesis suggests that scientists who do not win research awards will gradually lose their incentives for further research; (d) The age decrement hypothesis proposes that older scientists mostly conduct research with a lower intellectual and physical level than that of their younger colleagues; (e) The obsolescence hypothesis implies that the younger scientists use novel tools, techniques, and methodologies for research more easily than their older colleagues; (f) The intellectual deadlock hypothesis suggests that older scientists have less tendency to “reorient their research towards new scientific or social problems” (Kyvik and Olsen, 2008, pp. 442).

In terms of funding, Manjarrés-Henríquez et al. (2009) found that industrial R&D contracts and private funding are effective if the professor is in need of money. In the other words, "R&D contracts with industry and academic research activities have synergistic effects on scientific production, but only when R&D contracts account for a small percentage of a researcher's total funding, otherwise, there are decreasing marginal returns to scientific output" (pp. 799). Feldman and Graddy-Reed (2013) also showed some evidence about the positive effect of philanthropic funding coming from not-for-profit (NFP) organisations.

There are many examples for positive effect of public funding on scientific production. One famous justification for developing public support of science is that it is a public good (Partha and David, 1994). From another standpoint, Callon (1994) argued that Science is not a public good because of its intrinsic and natural properties, but a public good due to the fact that it is a source of diversity and flexibility. The fall in proportion of public funding relative to the other sources of funds has some influence on the university research trend. Geuna (2001) argued that the fading role of public funding can result in over-use of resources, focus on short-term research endeavour, conflict in incentive structures, and "exacerbation of the impact of cumulative and self-reinforcement phenomena present in the process of scientific production" (pp. 626). In other words, the abovementioned disadvantages can become positive when public funding increases. In terms of private funding, there are some articles questioning the positive and significant effect of private funding on scientific productivity, implying that private funding has either a neutral or a detrimental effect (Hottenrott and Thorwarth, 2011; Kleinman and Vallas, 2001).

Other university-specific effects also influence scientists' production. Crane (1965) indicated that scientists in major universities are more likely to be productive than scientists in minor universities. The main justification for such a phenomenon comes from the fact that more prestigious universities are better able to select the best students, are richer and hence can recruit better researchers, and are more apt at providing research procurement. Kyvik (1990), Blackburn et al. (1978), and Landry et al. (2007) also mentioned the variation of publication counts in different fields.

4.4 Data and methodology

4.4.1 Data and variables

In order to validate the hypothesis, a data set based on the integration of Quebec scientists' data measuring their funding and publications was built. In terms of publications, Thompson Reuters Web of Science provides information on scientific articles (date of publication, journal name, authors, coauthors, citations, and authors' affiliations). In addition, the Quebec University Research Information System (*Système d'information sur la recherche universitaire* or SIRU) of the Ministry of Education and Research was used to extract the grants and contracts, including yearly amounts, source, type, or other funding information of all Quebec university scientists during the period 2000-2010. These two databases were merged based on a scientist's unique identifier that was perfectly disambiguated by years of work at the Observatoire des sciences et des technologies (OST) in Montreal.

The dependent variable in our regression equations is an indicator of scientist's production and is measured by the natural logarithm of the yearly number of published articles [*ln(nbArticle)*].

The pertinent variables to validate our hypothesis identifies whether amongst the collaborator of an individual, there are co-authors who are in the top 10% and top 5% most funded individuals, distinguishing total funding, public sector funding, and private sector funding. To generate the variables measuring collaboration with top-funded scientists, a set of dummy variables has been generated to identify whether a scientist has collaborated with a top funded scientist: *ColT90*, and *ColT95* are the dummy variables that are equal to 1 if any of the coauthors in that year is amongst the top funded scientists (top 10%, and top 5% of total funding respectively). The funding amount used to generate these dummies is the sum of operational funding and funding for purchasing instruments. The variables of *ColPub90* and *ColPub95* are similarly built but only for public funding of coauthors and the variables of *ColPriv90* and *ColPriv95* have the same method of generation but for private funding of coauthors.

The database also includes other determinants of scientific productivity. Research funding can be awarded from different sources: the public sector, the private sector, or an organization with social and political missions, i.e. not-for-profit organisation (NFP). In addition, research funds can serve two purposes: it is directly used for research cost and researchers' salary (operational cost – O) or it indirectly helps research teams in buying instrument or logistic expenditure of laboratories (infrastructure costs – I). With these two mentioned categories and the three types of funding sources, it is possible to generate six research funding variables for each researcher [*ln(PublicfundingO)*, *ln(PublicfundingI)*, *ln(PrivatefundingO)*, *ln(PrivatefundingI)*, *ln(NFPfundingO)*, and *ln(NFPfundingI)*]. Although we have information for these six variables, this research only focuses on the effect of the operational budget because funding for the purpose of research tools and instruments does not have a regular pattern, i.e. it depends on the research needs, field, and handiness of updated research instruments. Hence infrastructure grants will not be used in this paper. To make the individual funding more informative, the amount of funds which is awarded to a team of researchers should be divided to number of researchers in the team. In addition, the funding variables are measured in three-year averages to smooth out large variations in yearly funding.

To complete the dataset, we add the yearly average number of authors in the papers of each scientist [*nbAuthor*]. In addition, individual characteristics regarding age and gender of scientist are also added to the data [*Age*, *dFemale*]. Description and summary of variables are available in the appendices section.

4.4.2 Methodology and econometrics model

To measure the effect of "collaboration with top-funded scientists" on a scientist's productivity (the number of articles), a regression equation is fitted to the available data. Because we have information on a number of years for many researchers, the data base is built as a panel. In addition to the dummy variables of collaboration with top-funded scientists as the main independent right-hand-side (RHS) variable, the left-hand-side (LHS) variable of the regression should also include control variables that affect the number of articles. Among others, gender, funding, and research field are important determinants.

Reviewing the literature in section 1, the RHS includes the dummy variables indicating whether the scientist has any collaboration with top-funded scientists [*ColT90, ColT95, ColPub90, ColPub95, ColPriv90, ColPriv95*], funding for operational cost [*ln(PublicfundingO), ln(PrivatefundingO), ln(NFPfundingO)*], gender [*dFemale*], age [*Age*] and dummies for research field, year, and universities.

One may argue that research collaboration increases scientific productivity because of the simple fact that writing an article as a single author may take double the time of writing an article with two authors. To control for this issue and for the obvious impact it may have on scientific production, we have added the lagged number of co-authors [*ln(nbAuthor)*] to the RHS variables.

In addition, we control for university and research field effects to account for any impact that our explanatory variables do not cover. For example, McGill University and University of Montreal (UdeM) produce more scientific publications (figure 4.1). The small universities are grouped according to their active disciplines and other institutional similarities. The University of Quebec and Bishop University are in the same group. The second group includes “École de technologie supérieure” (ETS), “Université du Québec à Montréal” (UQAM), and “Institut national de la recherche scientifique” (INRS). In terms of research field, the fields of Science, Engineering, Medical Science, and Health Science (figure 4.2) are more productive than others. We also add year dummy variables to account for year-specific characteristics of the research system as exemplified by the evolution of article counts over time (figure 4.3). The significant time trend and differences between different universities/research fields justify the existence of these dummy variables in the model.

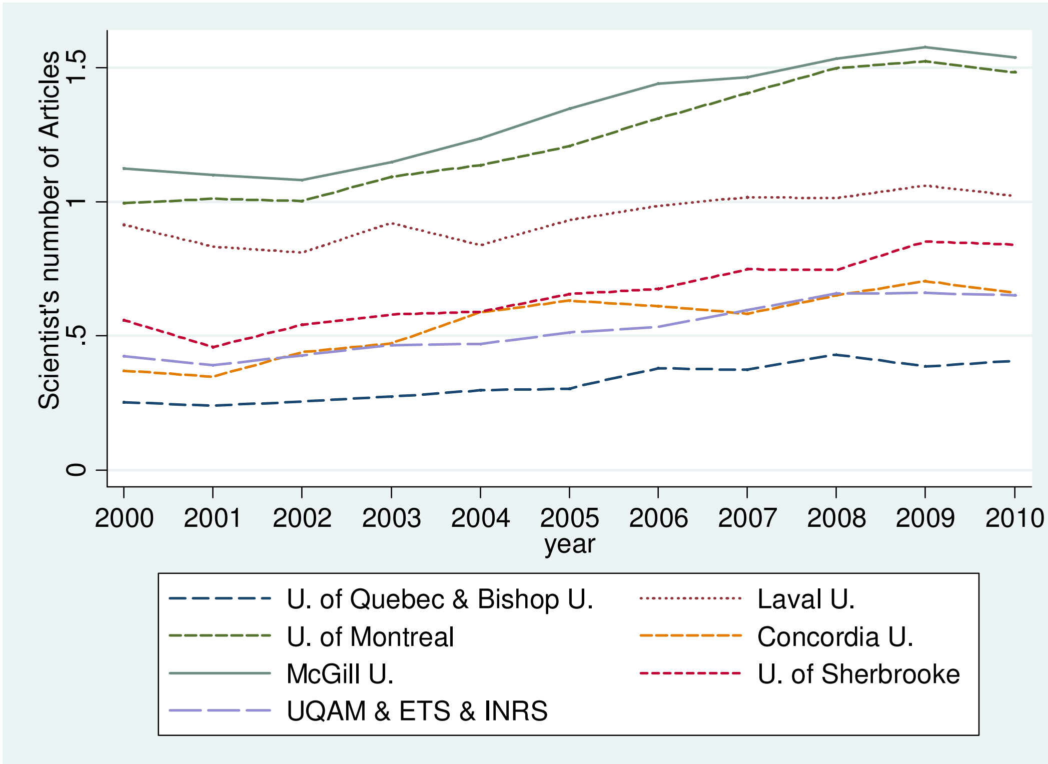

Figure 4.1: Number of articles published by the scientists of each university

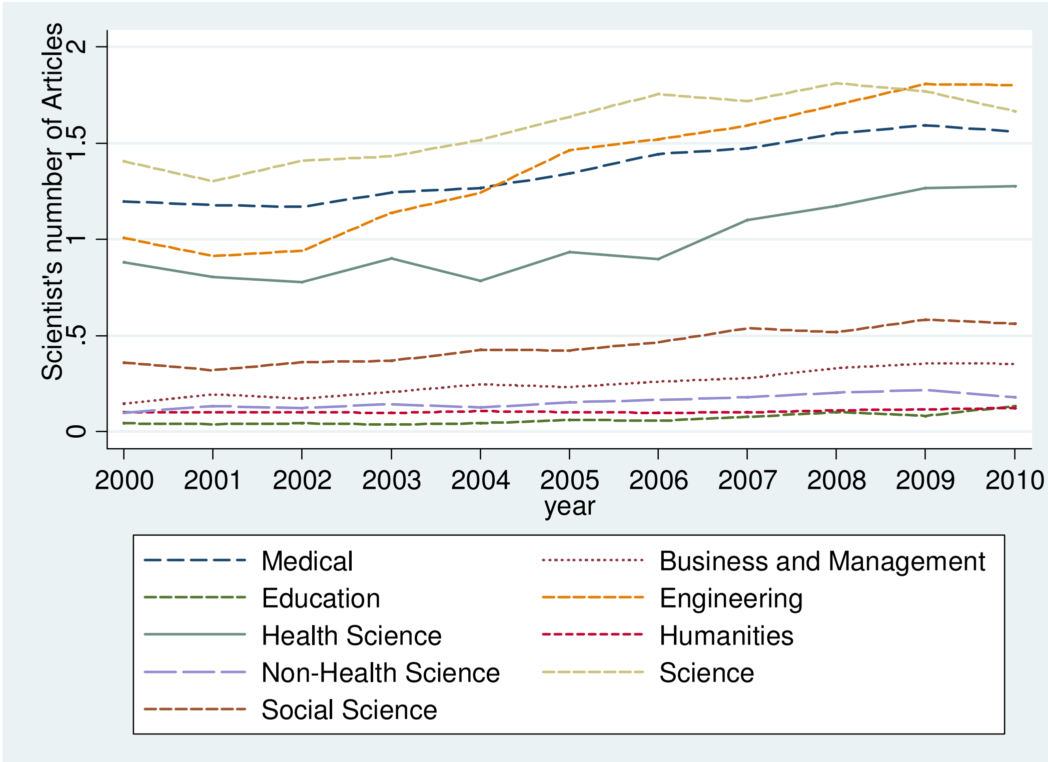

Figure 4.2: Number of articles published by the scientists in different fields

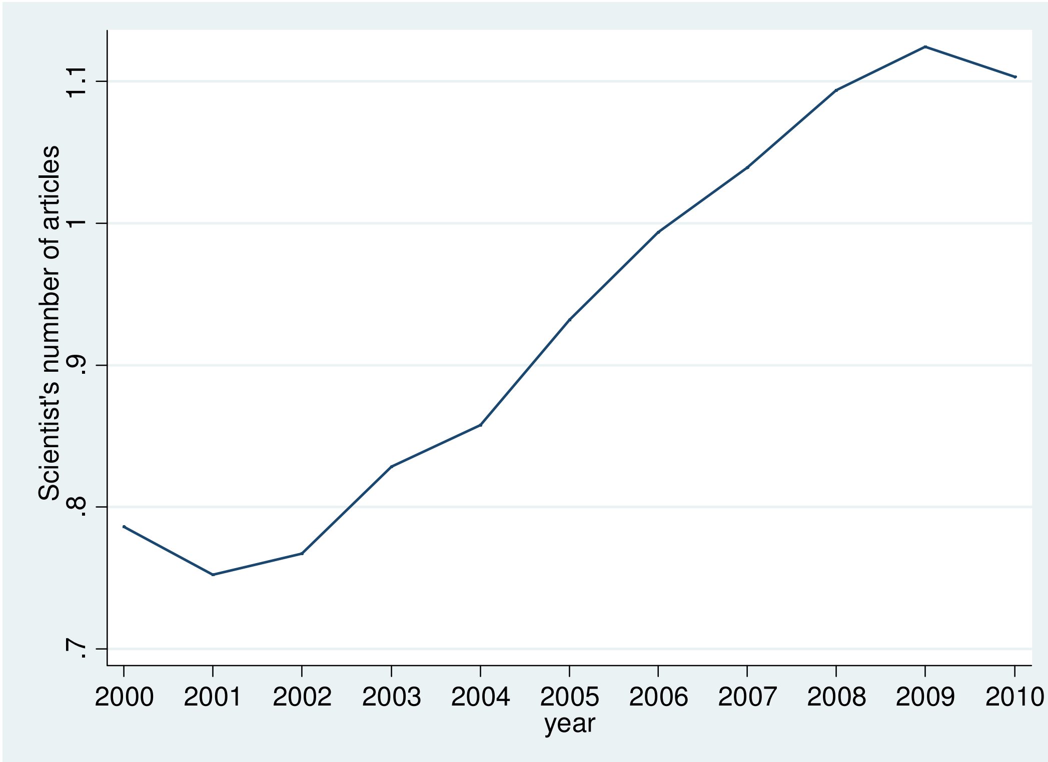

Figure 4.3: Average number of articles per scientist in each year

The possible reason behind yearly differences is that research fluctuates each year based on the economy and research policy and such fluctuation may affect research impact. University dummies and research field dummies can have the same role as academic publication norms and standards, research settings and related motivations are partially university and field dependent. The “Year 2000”, “McGill University”, and the field of “Medical science” are selected as reference point in dummy analysis. Considering the mentioned explanatory variables, the resulting model is given by:

$$\ln\left(nbArticle_{it}\right) = f\begin{pmatrix} \ln\left(PublicFundingO_{it}\right), \ln\left(PrivateFundingO_{it}\right), \ln\left(NFPFundingO_{it}\right), \\ ln\left(nbAuthor_{it}\right), ColTX_{it}, dFemale_{i}, Age_{it}, Age_{it}^{2}, D_{Field}, D_{University}, D_{Year} \end{pmatrix}$$

where *T* (in *ColTX*) is calculated either on the total amount of funding of the top-funded scientist, the total amount of public funding or the total amount of private funding, *X* (in *ColTX*) represents the top 75%, 90% or 95% most funded scientists, and the *D*s account for field, university and year dummy variables.

It is important to note that because the two variables [ln(*PublicfundingO*)] and [ln(*nbArticle*)] are determined by each other, which is the source of endogeneity, the ordinary least square (OLS) models are biased. The main reason for this potential endogeneity is that the scientists are assessed in their demands for public funding based on their CV and past effectiveness while at the same time, publication and research impact depend on the funding capability of researchers.

Using instrumental variables (IV) is a commonly suggested method to address endogeneity problems. There are two requirements for using IVs: (1) the instruments must be correlated with the endogenous variable, and (2) the instruments should not be correlated with the error term in the main regression equation. In other words, the instruments cannot suffer from the same endogeneity problem. If there is more than one instrument for the endogenous variable, it is necessary to perform a two-stage regression, in which the first stage estimates the endogenous variable (or instrumented variable) on a list of instrumental variables. Such estimation removes the error term of first stage and keeps the estimated amount for the second stage. Neglecting the error term of the endogenous variable and putting the estimated amount in the main regression equation is one way to remove the correlation of the RHS variables with the error term and to deal with the endogeneity problem.

A few variables are potential instruments for the amount of public funding. The number of scientists in a university [*nbScientistUni*] can explain the allocation of money amongst scientists. We anticipate that a university with a higher number of scientists may be able to benefit from cost sharing of research expenditures, hence reducing the need for larger amounts of individual funding. The rank of a scientist in the field is another variable that can be used to explain research funds allocation. There is an established tradition in the literature to compare past research productivity of scientists to allocate funds (Ho et al., 2006; Liefner, 2003).

The rank of previous funding can be another choice to predicting the future amount of funding. There is an echo effect for the amount of funding, which means that highly funded scientists are better able to get new sources of research money. The logic behind this argument is that decent research funds have effective networking capacity (Winter et al., 2006) to create different opportunities to get funds in a country such as Canada (Salazar and Holbrook, 2007). It should also be noted that the rank of funding is an ordinal variable and as such does not have information about amounts (the amount of research fund may indicate the capacity of knowledge production while the rank of funding does not provide such information). Therefore, it is an informative signal to estimate the amount of funding but it is not the funding itself. The rank of a scientist in the field in terms of three-year average of funding for the purpose of operational costs and direct expenditure of research [*PubORank*] is the second instrument to predict the amount of funding.

The third instrument is the total funding of each research cluster in each university [*totPublicfundingOcluster*]. Having known the aggregate amount of funding, the rank of scientist in terms of funding, and the number of university, it is possible to make an estimation for funding of each scientist. In the first stage, the amount of public funding [ln(*PublicfundingO*)] is estimated by the instruments and the variables of the second stage regression. To avoid simultaneity problems, public funding is not contemporaneous to the instruments; hence one-year lags of the instruments are used in the first-stage regression. Considering the mentioned explanatory variables, the resulting model is given by:

$$1^{st}\ stage: \ln\left(PublicFundingO_{it}\right) = g\left(PubORank_{it-1}, \ln(totPublicfundingOcluster_{it-1}), \ln\left(nbScientistUni_{it-1}\right)\right)$$

$$2^{nd}\ stage: \ln\left(nbArticle_{it}\right) = f\begin{pmatrix} \ln\left(PublicFundingO_{it}\right), \ln\left(PrivateFundingO_{it}\right), \ln\left(NFPFundingO_{it}\right), \\ \ln\left(nbAuthor_{it}\right), ColTX_{it}, dFemale_{i}, Age_{it}, Age_{it}^{\ 2}, D_{Field}, D_{University}, D_{Year} \end{pmatrix}$$

In a well-specified model, the variables of the RHS (including the instrumental variables) should not be highly correlated with each other. A low correlation refers to a good level of independence and explanatory power of RHS variables. The correlation matrix is reported in the appendices section and the correlation coefficients are acceptable for estimating regression equations.

4.5 Results and discussion

Because a number of our independent variables are individual effects (i.e. gender, university affiliation for vast majority of scientists), which are constant over ime, we prefer to estimate random effect 2SLS regressions for panel data (with the *xtivreg* command in Stata). The second stage of the two-stage panel regressions for collaboration with top 5% and 10% (in terms of public funding, private funding, and funding from not-for-profit sector) are presented in tables 4.1 to 4.5. The first stage regressions are not reported in the text to save space but the instruments significantly explain the public funding endogenous variable.

Our results indicate that the coefficients of the amounts of funding from the public, private, and not-for-profit sectors are all significant with positive signs in all regressions: the more funded the scientists are the more articles they publish. Having sufficient funds is a necessary condition for researchers to buy instruments and hire assistants in order to follow new ideas and conduct research at the frontier of knowledge. Our findings hence echo past studies in terms of public funding (Feldman and Graddy-Reed, 2013; Salter and Martin, 2001) and private funding (Manjarrés-Henríquez et al., 2009; Manjarrés-Henríquez et al., 2008).

In addition to the funding, our results show that the collaboration with top funded scientists is another significant determinant to foster the scientific productivity. There are different reasons for this effect. According to the papers assessing the effect of funding on scientific productivity (Beise and Stahl, 1999; Salter and Martin, 2001), it is possible to say that well-funded scientists have a better productivity. Collaboration with these scientists increases the chance of getting valuable experience and learning new skills through collaborative and collective research activities. The scientists who collaborated with well-funded scientists are more likely to publish papers.

Economies of scale are the second possible reason behind our results. For instance, with a greater number of students that produce more articles, the per-unit fixed cost of shared resources (for example the cost of laboratory technician) is lower, which then leads to a decreased marginal cost of publication. The same can be said of instruments and infrastructure. The third explanation stems from having access to modern instrumentation that more funds can provide. Neumann and Finaly-Neumann (1990) argued that research publication is directly influenced by support such as research instruments. Scientists who collaborate with top-funded scientists have access to better instruments and this would result in more scientific products.

Furthermore, the possible fourth reason is related to the fact that large amounts of funding are generally given to a team of researchers. It is important to note that SIRU accounts for interuniversity transfers. Hence a scientists who transfers funds to a colleague in another university will have his funds reduced by the transferred amount. The recipient will thus have an increased amount of funds as a consequence. It is however not possible to rack funds from the same grants across universities as the reporting is not standardized.

Regarding past networking, we can also argue that having higher a number of co-authors in the past significantly increases scientific production, hence supporting (Johnes, 1988; Melin, 1996).

Turning to the other independent variables in regression equations, we conclude that female scientists have a significantly lower scientific production in all regressions. Long (1990) explains that women's opportunities for collaboration are significantly less than those of men's because women have young children. Kyvik and Teigen (1996) find that childcare and lack of research collaboration are the two factors which negatively affect the scientific publishing of women. In another study, Leahey (2006a) argues that the reason of low production of women is that women specialize less than men, which is an important factor for research production.

The age of a scientist also exhibits a significant inverted U-shaped effect with a peak at age 51. In other words, scientists have increasing scientific productivity until they reach the age of 51 but their productivity drops afterward. This statement is compatible with earlier works that explain the age effect (Bernier et al., 1975; Diamond, 1986; Levin and Stephan, 1991). Our model also verifies the effect of university and research discipline in addition to the year-specific effect on scientific production.

Tables 4.2 to 4.5 present the regressions where "collaboration with top funded scientists" is interacted with "other determinants of scientific productivity". Table 4.2 shows that having had a greater number of co-authors in the past increases scientific productivity, a phenomenon that is amplified if the researcher has collaborated with top-funded scientists. Table 4.3 indicates that although female scientists are less productive than their male colleagues in general, collaborating with top-funded scientists partially alleviates this negative effect. Table 4.4 and 4.5 also show the interactive effect of private funding and non-for-profit funding when a scientist has research collaboration with top-funded scientist. For private funding, there are only three significant interactive effects (collaboration with top-funded scientists at 10% level). The interesting point is that if a scientist has research collaboration with another scientist who is top-funded in terms of private funding, the effect of own private funding becomes negative. The other two interactive effects are positive. For non-for-profit funding, the story is a bit different and interactive variables have positive and significant effect except for [*ColT95*ln(NFPfundingO)*] and [*ColPub95*ln(NFPfundingO)*], which are insignificant. As reviewed, these interactive effects do not have same pattern and significance and it is not possible to make a general statement for that.

Table 4.1: The second stage regression results

Dependent Var: nbArticle	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
dFemale$_i$	-0.0824 ***	-0.0827 ***	-0.0839 ***	-0.0836 ***	-0.0841 ***	-0.0841 ***
	0.0116	0.0115	0.0115	0.0115	0.0116	0.0117
Age$_{it}$	0.0328 ***	0.0331 ***	0.0318 ***	0.0325 ***	0.0339 ***	0.0343 ***
	0.0034	0.0034	0.0034	0.0034	0.0035	0.0035
Age$_{it}^2$	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ln(*nbAuthor$_{it}$*)	0.1645 ***	0.1695 ***	0.1676 ***	0.1739 ***	0.1742 ***	0.1774 ***
	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066
ln(*PublicfundingO$_{it}$*)	0.0445 ***	0.0467 ***	0.0431 ***	0.0459 ***	0.0500 ***	0.0503 ***
	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
ln(*PrivatefundingO$_{it}$*)	0.0059 ***	0.0059 ***	0.0065 ***	0.0066 ***	0.0025 ***	0.0043 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
ln(*NFPfundingO$_{it}$*)	0.0045 ***	0.0048 ***	0.0052 ***	0.0053 ***	0.0054 ***	0.0056 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
ColT90$_{it}$	0.1802 ***					
	0.0058					
ColT95$_{it}$		0.1819 ***				
		0.0062				
ColPub90$_{it}$			0.1760 ***			
			0.0056			
ColPub95$_{it}$				0.1723 ***		
				0.0061		
ColPriv90$_{it}$					0.1725 ***	
					0.0060	
ColPriv95$_{it}$						0.1641 ***
						0.0067
Constant	-0.2738 ***	-0.2828 ***	-0.2380 ***	-0.2662 ***	-0.3384 ***	-0.3387 ***
	0.0896	0.0896	0.0894	0.0894	0.0901	0.0903
Number of observations	32299	32299	32299	32299	32299	32299
Number of scientists	5602	5602	5602	5602	5602	5602
$\chi 2$	5389	5268	5428	5258	5104	4871
sigma	0.4820	0.4814	0.4809	0.4799	0.4836	0.4846
rho	0.4078	0.4034	0.4052	0.3982	0.4080	0.4065
R^2 within groups	0.0504	0.0462	0.0508	0.0443	0.0442	0.0384
R^2 overall	0.2667	0.2632	0.2688	0.2642	0.2569	0.2520
R^2 between groups	0.4423	0.4397	0.4454	0.4439	0.4323	0.4283

*, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

Table 4.2: The second stage regression results (joint effect of research collaboration and number of articles)

Dependent Var: nbArticle	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$dFemale_i$	-0.0821 ***	-0.0826 ***	-0.0837 ***	-0.0835 ***	-0.0838 ***	-0.0841 ***
	0.0116	0.0116	0.0115	0.0115	0.0116	0.0117
Age_{it}	0.0327 ***	0.0332 ***	0.0317 ***	0.0325 ***	0.0337 ***	0.0343 ***
	0.0034	0.0034	0.0034	0.0034	0.0035	0.0035
Age_{it}^2	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ln($nbAuthor_{it}$)	0.1523 ***	0.1645 ***	0.1513 ***	0.1647 ***	0.1607 ***	0.1769 ***
	0.0078	0.0073	0.0076	0.0072	0.0074	0.0071
ln($PublicfundingO_{it}$)	0.0447 ***	0.0468 ***	0.0433 ***	0.0460 ***	0.0503 ***	0.0503 ***
	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
ln($PrivatefundingO_{it}$)	0.0059 ***	0.0058 ***	0.0065 ***	0.0066 ***	0.0026 ***	0.0043 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
ln($NFPfundingO_{it}$)	0.0045 ***	0.0048 ***	0.0052 ***	0.0053 ***	0.0054 ***	0.0056 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
$ColT90_{it}$	0.1312 ***					
	0.0175					
$ColT95_{it}$		0.1525 ***				
		0.0199				
$ColPub90_{it}$			0.1080 ***			
			0.0171			
$ColPub95_{it}$				0.1154 ***		
				0.0194		
$ColPriv90_{it}$					0.0976 ***	
					0.0194	
$ColPriv95_{it}$						0.1602 ***
						0.0229
$ColT90_{it}$*ln($nbArticle_{it}$)	0.0293 ***					
	0.0099					
$ColT95_{it}$*ln($nbArticle_{it}$)		0.0167				
		0.0108				
$ColPub90_{it}$*ln($nbArticle_{it}$)			0.0407 ***			
			0.0097			
$ColPub95_{it}$*ln($nbArticle_{it}$)				0.0324 ***		
				0.0105		
$ColPriv90_{it}$*ln($nbArticle_{it}$)					0.0441 ***	
					0.0109	
$ColPriv95_{it}$*ln($nbArticle_{it}$)						0.0022
						0.0123
$ColT90_{it}$*$dFemale_i$						
Constant	-0.2564 ***	-0.2783 ***	-0.2142 ***	-0.2543 ***	-0.3188 ***	-0.3382 ***
	0.0898	0.0897	0.0896	0.0896	0.0902	0.0904
Number of observations	32299	32299	32299	32299	32299	32299
Number of scientists	5602	5602	5602	5602	5602	5602
$\chi 2$	5399	5262	5455	5264	5123	4869
sigma	0.4820	0.4818	0.4808	0.4801	0.4834	0.4847
rho	0.4083	0.4044	0.4053	0.3989	0.4082	0.4068
R^2 ***within groups***	0.0507	0.0463	0.0514	0.0446	0.0448	0.0384
R^2 ***overall***	0.2663	0.2630	0.2686	0.2641	0.2567	0.2519
R^2 ***between groups***	0.4420	0.4393	0.4454	0.4436	0.4316	0.4282

*, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

Table 4.3: The second stage regression results (joint effect of research collaboration and scientist's gender)

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
$dFemale_i$	-0.0915 ***	-0.0909 ***	-0.0817 ***	-0.0898 ***	-0.0904 ***	-0.0912 ***
	0.0123	0.0119	0.0122	0.0118	0.0121	0.0119
Age_{it}	0.0328 ***	0.0331 ***	0.0317 ***	0.0325 ***	0.0340 ***	0.0344 ***
	0.0034	0.0034	0.0034	0.0034	0.0035	0.0035
Age_{it}^2	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\ln(nbAuthor_{it})$	0.1648 ***	0.1697 ***	0.1679 ***	0.1741 ***	0.1742 ***	0.1773 ***
	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066
$\ln(PublicfundingO_{it})$	0.0446 ***	0.0468 ***	0.0431 ***	0.0460 ***	0.0501 ***	0.0503 ***
	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
$\ln(PrivatefundingO_{it})$	0.0059 ***	0.0059 ***	0.0065 ***	0.0066 ***	0.0025 ***	0.0043 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
$\ln(NFPfundingO_{it})$	0.0045 ***	0.0048 ***	0.0052 ***	0.0053 ***	0.0054 ***	0.0056 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
$ColT90_{it}$	0.1734 ***					
	0.0065					
$ColT95_{it}$		0.1727 ***				
		0.0070				
$ColPub90_{it}$			0.1778 ***			
			0.0064			
$ColPub95_{it}$				0.1653 ***		
				0.0069		
$ColPriv90_{it}$					0.1668 ***	
					0.0067	
$ColPriv95_{it}$						0.1542 ***
						0.0075
$ColT90_{it}*dFemale_i$	0.0287 ***					
	0.0129					
$ColT95_{it}*dFemale_i$		0.0403 ***				
		0.0142				
$ColPub90_{it}*dFemale_i$			-0.0070			
			0.0127			
$ColPub95_{it}*dFemale_i$				0.0304 **		
				0.0141		
$ColPriv90_{it}*dFemale_i$					0.0257 *	
					0.0137	
$ColPriv95_{it}*dFemale_i$						0.0460 ***
						0.0157
Constant	-0.2735 ***	-0.2816 ***	-0.2372 ***	-0.2656 ***	-0.3394 ***	-0.3403 ***
	0.0895	0.0896	0.0894	0.0894	0.0901	0.0903
Number of observations	32299	32299	32299	32299	32299	32299
Number of scientists	5602	5602	5602	5602	5602	5602
$\chi 2$	5401	5285	5440	5274	5109	4884
sigma	0.4815	0.4809	0.4804	0.4793	0.4835	0.4844
rho	0.4070	0.4025	0.4039	0.3970	0.4079	0.4063
R^2 within groups	0.0506	0.0466	0.0507	0.0445	0.0444	0.0388
R^2 overall	0.2662	0.2626	0.2690	0.2638	0.2567	0.2516
R^2 between groups	0.4414	0.4386	0.4457	0.4430	0.4319	0.4276

*, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

Table 4.4: The second stage regression results (joint effect of research collaboration and private funding)

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
dFemale$_i$	-0.0824 ***	-0.0827 ***	-0.0841 ***	-0.0836 ***	-0.0844 ***	-0.0842 ***
	0.0115	0.0115	0.0114	0.0114	0.0116	0.0117
Age$_{it}$	0.0326 ***	0.0330 ***	0.0315 ***	0.0323 ***	0.0339 ***	0.0343 ***
	0.0034	0.0034	0.0034	0.0034	0.0035	0.0035
Age$_{it}^2$	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ln(*nbAuthor$_{it}$*)	0.1655 ***	0.1699 ***	0.1695 ***	0.1746 ***	0.1734 ***	0.1771 ***
	0.0066	0.0066	0.0066	0.0066	0.0066	0.0067
ln(*PublicfundingO$_{it}$*)	0.0446 ***	0.0468 ***	0.0432 ***	0.0459 ***	0.0500 ***	0.0503 ***
	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
ln(*PrivatefundingO$_{it}$*)	0.0048 ***	0.0059 ***	0.0046 ***	0.0063 ***	0.0045 ***	0.0048 ***
	0.0009	0.0008	0.0009	0.0008	0.0009	0.0008
ln(*NFPfundingO$_{it}$*)	0.0045 ***	0.0048 ***	0.0052 ***	0.0053 ***	0.0054 ***	0.0056 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
ColT90$_{it}$	0.1725 ***					
	0.0070					
ColT95$_{it}$		0.1819 ***				
		0.0078				
ColPub90$_{it}$			0.1618 ***			
			0.0069			
ColPub95$_{it}$				0.1685 ***		
				0.0076		
ColPriv90$_{it}$					0.1907 ***	
					0.0077	
ColPriv95$_{it}$						0.1740 ***
						0.0093
ColT90$_{it}$*ln(PrivatefundingO$_{it}$)	0.0022 **					
	0.0011					
ColT95$_{it}$*ln(PrivatefundingO$_{it}$)		0.0000				
		0.0011				
ColPub90$_{it}$*ln(PrivatefundingO$_{it}$)			0.0039 ***			
			0.0011			
ColPub95$_{it}$*ln(PrivatefundingO$_{it}$)				0.0010		
				0.0011		
ColPriv90$_{it}$*ln(PrivatefundingO$_{it}$)					-0.0043 ***	
					0.0012	
ColPriv95$_{it}$*ln(PrivatefundingO$_{it}$)						-0.0019
						0.0012
Constant	-0.2693 ***	-0.2811 ***	-0.2284 ***	-0.2624 ***	-0.3422 ***	-0.3397 ***
	0.0894	0.0895	0.0891	0.0893	0.0901	0.0903
Number of observations	32299	32299	32299	32299	32299	32299
Number of scientists	5602	5602	5602	5602	5602	5602
$\chi 2$	5428	5290	5506	5291	5118	4872
sigma	0.4805	0.4805	0.4783	0.4786	0.4834	0.4846
rho	0.4042	0.4010	0.3987	0.3948	0.4080	0.4066
R^2 within groups	0.0503	0.0462	0.0508	0.0442	0.0447	0.0385
R^2 overall	0.2674	0.2633	0.2704	0.2646	0.2567	0.2519
R^2 between groups	0.4435	0.4398	0.4478	0.4444	0.4317	0.4281

*, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

Table 4.5: The second stage regression results (joint effect of research collaboration and NFP funding)

Dependent Var: nbArticle	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$dFemale_i$	-0.0823 ***	-0.0827 ***	-0.0839 ***	-0.0836 ***	-0.0843 ***	-0.0843 ***
	0.0115	0.0115	0.0114	0.0114	0.0116	0.0116
Age_{it}	0.0325 ***	0.0329 ***	0.0315 ***	0.0322 ***	0.0337 ***	0.0342 ***
	0.0034	0.0034	0.0034	0.0034	0.0034	0.0035
Age_{it}^2	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ln(*$nbAuthor_{it}$*)	0.1655 ***	0.1702 ***	0.1691 ***	0.1749 ***	0.1753 ***	0.1783 ***
	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066
ln(*$PublicfundingO_{it}$*)	0.0446 ***	0.0468 ***	0.0432 ***	0.0460 ***	0.0502 ***	0.0504 ***
	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
ln(*$PrivatefundingO_{it}$*)	0.0059 ***	0.0059 ***	0.0065 ***	0.0066 ***	0.0024 ***	0.0043 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
ln(*$NFPfundingO_{it}$*)	0.0035 ***	0.0045 ***	0.0039 ***	0.0048 ***	0.0038 ***	0.0048 ***
	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007
$ColT90_{it}$	0.1720 ***					
	0.0072					
$ColT95_{it}$		0.1769 ***				
		0.0081				
$ColPub90_{it}$			0.1645 ***			
			0.0071			
$ColPub95_{it}$				0.1655 ***		
				0.0080		
$ColPriv90_{it}$					0.1546 ***	
					0.0076	
$ColPriv95_{it}$						0.1480 ***
						0.0089
$ColT90_{it}$*ln($NFPfundingO_{it}$)	0.0020 **					
	0.0011					
$ColT95_{it}$*ln($NFPfundingO_{it}$)		0.0011				
		0.0011				
$ColPub90_{it}$*ln($NFPfundingO_{it}$)			0.0029 ***			
			0.0010			
$ColPub95_{it}$*ln($NFPfundingO_{it}$)				0.0015		
				0.0011		
$ColPriv90_{it}$*ln($NFPfundingO_{it}$)					0.0041 ***	
					0.0011	
$ColPriv95_{it}$*ln($NFPfundingO_{it}$)						0.0033 ***
						0.0012
Constant	-0.2680 ***	-0.2793 ***	-0.2308 ***	-0.2617 ***	-0.3325 ***	-0.3359 ***
	0.0894	0.0895	0.0891	0.0892	0.0899	0.0902
Number of observations	32299	32299	32299	32299	32299	32299
Number of scientists	5602	5602	5602	5602	5602	5602
$\chi 2$	5428	5295	5492	5302	5156	4906
sigma	0.4805	0.4804	0.4786	0.4782	0.4820	0.4836
rho	0.4042	0.4007	0.3994	0.3939	0.4043	0.4039
R^2 within groups	0.0503	0.0461	0.0507	0.0442	0.0443	0.0384
R^2 overall	0.2673	0.2636	0.2698	0.2647	0.2580	0.2527
R^2 between groups	0.4436	0.4403	0.4472	0.4448	0.4341	0.4296

*, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

4.6 Conclusion

In this article, a theoretical model was developed to show that collaboration with top-funded scientists has a positive effect on the number of articles published by a scientist. In addition, the empirical results from a two-stage panel regression verify the results predicted by the theoretical model. Using a rich dataset on publication and funding of Quebec scientists, the paper shows that collaborating with top-funded scientists has positive effect on the productivity of a scientist.

The results also verified the evidence from the literature regarding the positive effect of funding, the positive effect of past networking (measured by number of co-authors), the inverted U-shaped effect of age, and the fewer number of publications by women compared to men. We have also significant effect of university, research field, and year. This implies that each university, each research field, or each year has unmeasurable or hidden set of scientific, social, and financial characteristics affecting scientific productivity.

Considering the significant and positive effect of collaboration with top-funded scientists on scientific productivity, the main policy advice is to set some incentives or to give a mandate to top-funded scientists to have specific amount/number of collaboration with low-funded scientists. This kind of collaboration may positively affect the academic productivity through different channels. As discussed above it may result in a substantial transfer of tacit knowledge and more scientific publications. Moreover, it provides benefits such as economy of scale in knowledge production because well-funded scientists have generally larger teams of researchers and better research equipment. The expansion of the research network is another reviewed benefit of such collaboration. Collaboration with top-funded scientists has also some interactive effects with other determinants of scientific production, which suggests an amplified positive impact if a researcher has greater number of co-authors, if a researcher is female. In terms of interaction with private finding and NFP funding, there is not a systemic and structured pattern. However, some of interactive variables have positive effect.

In terms of validity of abovementioned interpretations, one should be careful as there are a number of limitations in this study. First this study only covers Quebec scientists. Second the database time coverage is limited. Third some data entries are missing in the original dataset and we do not have the funding information of all collaborators, only of Quebec collaborators. In addition to addressing these shortcomings in future studies with having more data entries and complete records, there are also some other suggestions for future works on this subject. First, this paper defines the "star" scientists based on the amount of funding, i.e. it identifies the top-funded scientists. As the funding of scientist is being determined by previous productivity and intrinsic characteristics of scientists, it is an appropriate proxy for research impact. However, future research can consider different definitions of "star scientist" to provide a more comprehensive and robust interpretation. For instance, star scientists can be identified based on the number of patents, the impact factor of the journals in which the researchers publish, and networking measure such as centrality or cliquishness indicators.

The second suggestion for future studies is about different research methods. One can have a deep investigation on the nature of research collaboration and provide a taxonomy to explain the benefit and knowledge spillover of different types of research collaboration with star scientists. This will require data gathering by surveys and questionnaires in order to have a detailed understating the nature of the collaboration with star scientists.

Chapter 5 **ARTICLE 3 : THE EFFECT OF HOLDING A RESEARCH CHAIR ON SCIENTISTS' PRODUCTIVITY**

Keywords: Scientist Publication, Research Funding, Matching Techniques, Research Chair

5.1 Abstract

Having combined data on Quebec scientists' funding and journal publication, this paper tests the effect of holding a research chair on a scientist's productivity. The novelty of this paper is to use a matching technique to understand whether holding a research chair contributes to a better scientific productivity. This method compares two different sets of regressions which are conducted on different data sets: one with all observations and another with only the observations of the matched scientists. Two chair and non-chair scientists are deemed matched with each other when they have the closest propensity score in terms of gender, research field, and amount of funding. The results show that holding a research chair is a significant scientific productivity determinant in the complete data set. However, when only matched scientists are kept in data set, holding a Canada research chair has a significant positive effect on scientific productivity but other types of chairs do not have a significant effect. In the other words, in the case of two similar scientists in terms of gender, research funding, and research field, only holding a Canada research chair significantly affects scientific productivity.

5.2 Introduction

Scientists' academic productivity has been extensively discussed and many of its determinants are currently known as potential motives for publishing papers in peer-reviewed journals. Among others, age, gender, private and public funding, institutional setting, field and context are the most important determinants. Funding definitely plays a major role in knowledge production and in shaping scientific productivity: its positive effect on scientific productivity has been extensively demonstrated in the literature (Crespi and Geuna, 2008; Pavitt, 2000, 2001; Salter and Martin, 2001).

The name and prestige of research centers may generally affect scientific productivity. For instance, Turner (2005) showed the outstanding research productivity of 'Grandes Ecoles' in France. Beside the name and brand of academic institutions, centers with specific research orientations such as 'centers for excellence' are also effective on the level of scientific productivity (Niosi, 2002).

Other desirable factors, similar to 'centers for excellence', may increase an individual's research motivation and influence the willingness or ability of a scientist for conducting original research. In this paper, we focus on the effect of holding a 'research chair' as a possible determinant of scientific publication. On the one hand, it may help the holder of this chair to be liberated from the constant quest for research funds or to have time to construct a more effective network, which may result in propelling future knowledge production. On the other hand, greater scientific productivity may simply be the effect of the past productivity of a scientist, implying an intrinsic ability of scientists in conducting research and/or in mobilising effectively its extensive networking capacity.

Using an appropriate econometric model, this paper tests whether 'holding a chair' has a significant effect on scientific productivity. The rest of paper is organised as follows: section 1 reviews the literature and proposes a theoretical framework that leads to two hypotheses. Section 2 explains how data is gathered and describes the variables used in the econometric models. Section 3 presents the results of the regressions in regard to testing the hypotheses. Finally, section 4 concludes and summarizes the results of the paper.

5.3 Theoretical framework

The literature relevant to this article brushes on the importance of having a prestigious academic position or affiliation. Focusing on the role of university prestige in academic productivity, Long et al. (1979) found a positive and significant correlation between the prestige of the scientist alma matter and the prestige of subsequent employment affiliation. The authors also indicated that graduating from a prestigious university has a positive effect on citations (but not on publication counts). The paper also provides a justification for the effect of prestige arguing that the best students are admitted to the most prestigious universities and subsequently the graduates of the prestigious universities are generally recruited by other similar institutions. Furthermore, such scientists who studied in and have been recruited by prestigious universities are better able to interact with new gifted students (Long et al., 1979). This paper argued that academic prestige can push forward research and its impact. More recently, Zhou et al. (2012) showed that papers cited by prestigious scientists, regardless of the number of citations, are of a higher impact than papers which are cited by 'ordinary' scientists.

Prestige can be seen from the reverse direction of causality. West et al. (1998) investigated the relationship between departmental climate, such as degree of formalization, support for career development and support for innovation on the one hand, and officially rated effectiveness of universities on the other hand. They conclude that the causality direction is from former to latter, showing that the prestige of universities is an effect and not a cause for appropriate departmental climate and necessary institutional setting for conducting research.

Measuring and quantifying prestige have been discussed in some articles. Frey and Rost (2010) compared three types of university ranking based on the number of articles, number of citations, and membership of editorial board or of academic associations. The authors indicated that these rankings are not compatible with each other and suggested the use of multiple measurements. Van Raan (2005) criticized the applicability of university rankings such as the Shanghai ranking for evaluating academic excellence by noting that the 'affiliation', as an important factor reflecting research atmosphere, is not well addressed in those ranking. In addition to the university ranking, it is important to assess individual research productivity to have a better sense of prestige. Henrekson and Waldenström (2007) introduced three types of indicators, measuring research productivity: (1) measures based on weighted journal publications, (2) measures based on citations to most cited works, and (3) measures based on the number of publications.

To measure prestige with a more robust measure, it is possible to consider awards as a measure of prestige, which is awarded based on a deliberate assessment in specialized and independent committees. Different types of research chairs are examples of awards. In Canada, there are three types of research chair: (1) research chairs which are awarded by industry and generally referred to as industrial chairs; (2) research chairs which are awarded by Canadian federal funding agencies such as the Natural Science and Engineering Research Council (NSERC) and the Canadian Institutes of Health Research (CIHR); and (3) the 'Canada research chairs', whose holders are assumed to already achieve research excellence in one of the main fields of research: engineering and the natural sciences, health sciences, humanities, and social sciences. The purpose of this latter program is to "improve […] depth of knowledge and quality of life, strengthen Canada's international competitiveness, and help train the next generation of highly skilled people through student supervision, teaching, and the coordination of other researchers' work. In terms of allocation, three separate calculations are performed for each eligible institution to determine their chair allocation within the three areas of research. The allocation method for regular chairs pools together each granting agency's funding for all universities and allocates chairs based on the portion of the granting agency's funding that each eligible institution has received. The funding received by each eligible university over three years is totalled. The portion of granting agency support that each eligible institution holds in this grand total determines the number of chairs allocated (i.e., the percentage of funding secured = the percentage of chairs allocated)."[26]. Considering holding a chair as some kind of measure of prestige, we aim to elucidate the effect of being a 'chair-holder' on scientific productivity. Our first hypothesis therefore reads as:

Hypothesis 1: *Holding a chair increases a scientist's productivity measured in terms of number of publications.*

Hypothesis 1 tests the productivity of chair-holders compared to other scientists and as such does not seek to prove causality. Considering the fact that chair-holders are the well-funded scientists too, this hypothesis cannot detach the funding effect of the chair from its other effects (mainly from

[26] http://www.chairs-chaires.gc.ca/about_us-a_notre_sujet/index-eng.aspx

prestige and networking effect). In other words, despite the evidence in literature about the benefits and goals of the research chair programs other than funding, hypothesis 1 is not able to disentangle them.

A number of authors tried to highlight the functions and characteristics of research chairs. Cantu et al. (2009) showed the research chair program would be a good strategy for implementing knowledge-based development. In study on German universities, Schimank (2005) argued that chair-holders are small businessmen with high job security and no bankruptcy in addition to a good level of freedom of teaching and research, indicating that research chairs have the characteristics of job security and sovereignty.

According to official documents, influencing scientific productivity is not the direct goal of a research chair. In the tenth-year evaluation report for Canada research chair (CRC)[27] programme, the authors conclude that CRC program is an effective way for Canadian universities to "attract and retain leading researchers" from other countries (page 4). The report does not say that holding a research chair is a determining factor of a chair's scientific production: "the extent to which this success can be related directly to the CRC is difficult to quantify" (page 5). Furthermore, holding a research chair does not imply a higher salary. Courty and Sim (2012) showed that although having Canada Research Chair (CRC) initially increases the professors' salary, such an increase erodes quickly over the time. This means that getting a research chair does not necessarily result in long-term salary jump.

It is possible to look at the research chair as a result of scientists' characteristics and achievements (for instance number of articles, and number of citations). A chair-holder may experience an effect specific to holding chair on his/her scientific production in addition to the contribution of his/her characteristics and past achievements. To disentangle the exclusive effect of chair from the effect of scientists' characteristics, we propose our second hypothesis:

Hypothesis 2: *Keeping a scientist's main characteristics (gender, research field, and amount of grants) constant, holding a research chair does not have a significant positive effect on his/her scientific production.*

[27] http://www.chairs-chaires.gc.ca/about_us-a_notre_sujet/publications/ten_year_evaluation_e.pdf

This hypothesis can be tested using a matching technique, which will be explained in the methodology section. In addition to our two hypotheses, a number of factors have been shown in the literature to influence research productivity. These will be used as control variables in our regression models. Among others, age, gender, funding, research field, and university characteristics are the most important determinants of scientific production which should be controlled when the effect of research chair on scientific productivity is being tested.

In terms of age, two strands of the literature about its effect on scientific productivity are in opposition to one another. First, some authors argue for the life cycle trend in economic activity, referring to the non-linearity of human productivity during life (Becker, 1962). Those articles show that productivity follows an inverted U-shape format (Kyvik, 1990; Kyvik and Olsen, 2008), which can be justified by the optimization of the trade-off between cost of human capital investment at a younger age and its return as a benefit at an older age. For instance, Bernier et al. (1975) showed an increase in impact and quantity of publications until the age of 44 and a fall after that age. The second strand of the literature generally finds that scientists' academic productivity (number of articles and number of citations) decreases with age (Bonaccorsi and Daraio, 2003; Diamond, 1986; Levin and Stephan, 1991).

Gender is also known as a significant determinant of scientific productivity in the literature. Nakhaie (2002) showed that Canadian female scientists have a lower scientific productivity compared to their male colleagues. Long (1990) explained that women's opportunities for collaboration are significantly less than those of men's because women have young children. However, in another study, Long (1992) showed that women are less productive in the first decade of their career but are more productive afterwards. Moreover, Xie and Shauman (1998) and Abramo et al. (2009a) argue that gender differences in research productivity have declined over time, while at the same time the population of female scientists has proportionally increased.

Research funding is another important determinant of scientific productivity. In a review article on the effect of funding, Salter and Martin (2001) suggested the following six types of contribution of publicly funded research: source of new useful knowledge, new instrumentation and methodologies, skills developed by those involved in carrying out basic research, expansion of national and international networks, dealing with complex problems[28], and creation of spin-off companies. Pavitt (2001) also referred to the importance of public support for scientific infrastructure development and highlights its role in the effectiveness of public grants. In another study, Pavitt (2000) argued that funding for infrastructure of expertise, equipment and networks is necessary for the development and implementation of research.

Another body of literature investigates the effect of university characteristics on scientific productivity. Heinze et al. (2009) showed that the small size of a research group, sufficient access to the various technical skills, and an appropriate leadership all result in an improvement in research productivity. Similarly, Carayol and Matt (2006) found that in a smaller laboratory, individual researchers publish more, compared with individuals working in a large laboratory. Other authors focused on the effect of faculty size. Buchmueller et al. (1999) indicated that graduate school faculty size is a significant determinant of the research proficiency of graduates. Jordan et al. (1988, 1989) suggested that research productivity is positively associated with department size but that the effect becomes weaker as the size increases. In contrast, Kyvik (1995) rejected both hypotheses that large departments are more productive and that faculty members of large departments better assess the research environment.

In terms of university governance, Jordan et al. (1989) found significant evidence that private institutions have a greater average productivity, but Golden and Carstensen (1992) found no difference between public and private universities in terms of research productivity when controlling for research support from leading research foundations and department faculty rating. Golden and Carstensen argued that public and private institutions are solely different in a way that public institutions have a greater teaching load, which may affect scientific productivity.

Differences between fields and context are also noted by a number of authors. Blackburn et al. (1978) showed that the fields of humanities and sciences have different patterns of scientific production. To justify the differences between disciplines, Baird (1986) for instance showed that large research laboratories in chemistry, the scholarly apprenticeship approach in history, and research over practice in psychology are important field-dependent factors in scientists' productivity. In another comprehensive study, Baird (1991) referred to the productivity and citation

[28] The paper argues that solving the complex problem provides great benefit for the firms and organizations facing such problems.

pattern differences among disciplines and argues that size, internal university support and federal support can explain such differences. The evidence from the literature clearly suggests that scientific productivity may depend on academic prestige and on other control variables such as funding, gender, age, and university-specific characteristics.

5.4 Section 2 - Data and methodology

5.4.1 Data and variables

In order to validate our two hypotheses, we built a data set based on the integration of data on funding and journal publications for Quebec scientists. For publications, Thompson Reuters Web of Science provides information on scientific articles (date of publication, journal name, authors and their affiliations). The dependent variable of our model therefore counts the yearly number of articles [*nbArticle*] published by an individual researcher in any given year. For each publication, the database also provides the number of co-authors. To control for team size, we therefore calculate the average number of co-authors [*nbAuthors*] over all the articles published by an individual researcher in any given year.

In terms of funding, we use a database of Quebec university researchers (*Système d'information sur la recherche universitaire* or SIRU) gathered and combined by the Ministry of Education and Research. This database lists the grants and contracts information, including yearly amount, source, and type for the period of 2000-2012 for all Quebec university scientists[29] and the title of each specific research project for which funding was granted. The titles of research project are being used to generate dummy variables identifying whether a scientist has a research chair; the title field clearly states: "chair in…". As mentioned above, three types of chair are available in Canada: (1) industrial research chairs [*dIndChair*]; (2) research chairs awarded by Canadian federal granting councils [*dGCChair*]; and (3) Canada research chairs [*dCRC*]. In addition, we created a dummy variable [*dIndGCChair*] indicating whether the scientist is an industrial chair or a chair assigned by Canadian federal granting councils (the combination of *dIndChair* and *dGCChair*). Finally, the

[29] When the funding is attributed to more than one recipient researcher, the total amount of funding is divided by the number of researchers in the team within the same university. The SIRU data accounts for all interuniversity transfers and funds are counted where they have been transferred and spent. Unfortunately, we have no means by which to sum the funds from the same grants that are transferred to other universities, as the reporting does not allow a match between the data.

dummy variable [*dChair*] is equal to 1 for scientists with any type of chair (the combination of *dIndChair*, *dGCChair, and dCRC*). Description and summary of variables are available in the appendices section.

The next set of variables in the data set measures funding information. In terms of source, the research funding can be awarded by the public sector, the private sector, or organizations with social and political missions, which we classify as not-for-profit (NFP) organisations. Our research will concentrate on operational cost[30] research funding to ensure that the three fields[31] examined are more comparable – there is relatively little infrastructure investment in the humanities and social sciences compared to health sciences and the engineering and natural sciences. The three funding variables to be considered are therefore [*PublicfundingO*, *PrivatefundingO*, and *NFPfundingO*] for the public, private and not-for-profits sectors respectively. The funding variables are measured in three-year averages to smooth out large variations in yearly funding.

In addition, age and gender of scientist are also available in dataset [*Age*, *dFemale*] and provide useful controls for scientific productivity, as was highlighted in the theoretical framework. This information was disambiguated by the Observatoire des sciences et des technologies (OST) for all Quebec academics.

5.4.2 Methodology and econometrics model

To measure the effect of 'holding a research chair' on a scientist's productivity, we use a panel regression model where the left-hand-side (LHS) variable the number of articles [*nbArticle*] is a measure of scientific productivity. On the right-hand-side (RHS), the main independent variables are the research chair dummy variables described above [*dIndChair, dGCChair, dCRC, dIndGCChair, dChair*]. The other independent variables are our controls, among others age [*Age*], gender [*dFemale*], as well as funding, and are described in the next paragraphs.

[30] Research funds may serve two purposes: they may be directly used for research cost and researchers' salary as operational costs (O) or indirectly help research teams in buying instruments or laboratory infrastructure (I). It is therefore possible to generate six research funding variables for each researcher [*PublicfundingO, PublicfundingI, PrivatefundingO, PrivatefundingI, NFPfundingO,* and *NFPfundingI*]. In reality, research infrastructure funding stems mainly from public sources and the private and not-for-profit sources (*PrivatefundingI, NFPfundingI*) are too sporadic, i.e. rarely different from 0, to be used effectively in our models.

[31] We have three fields: 'engineering and the natural sciences', 'health sciences', and' humanities, and social sciences'

In terms of funding, this research only focuses on the effect of the operational budget because funding for the purpose of buying instruments or laboratory infrastructure does not have a regular pattern, implying that it depends on the research needs, field, and handiness of updated research instruments. Hence we only use the variables of operational costs [*PublicfundingO, PrivatefundingO, NFPfundingO*] to control for the effect of funding. We also measure the interactive effect of funding and holding a chair on scientific productivity in regression models to find out whether there is a difference between the funding effect of chair-holders and of non-chair-holders.

Age [*Age*] and its square are inserted in regression to investigate a non-linear effect of age on scientific productivity. We also control for university, year, and research division effect in order to account for any impact that our explanatory variables may not cover. For example, McGill University and University of Montreal (UdeM) produce more scientific publications (figure 5.1[32]). In terms of research divisions, Science, Engineering, Medical Science, and Health Science (figure 5.2) are more productive than others. We also add year dummy variables to account for year-specific characteristics of the research system as exemplified by the evolution of article counts over time (figure 5.3).

[32] The small universities are grouped according to their active disciplines and other institutional similarities. The University of Quebec and Bishop University are in the same group. The second group includes "École de technologie supérieure" (ETS), "Université du Québec à Montréal" (UQAM), and Institut national de la recherche scientifique (INRS).

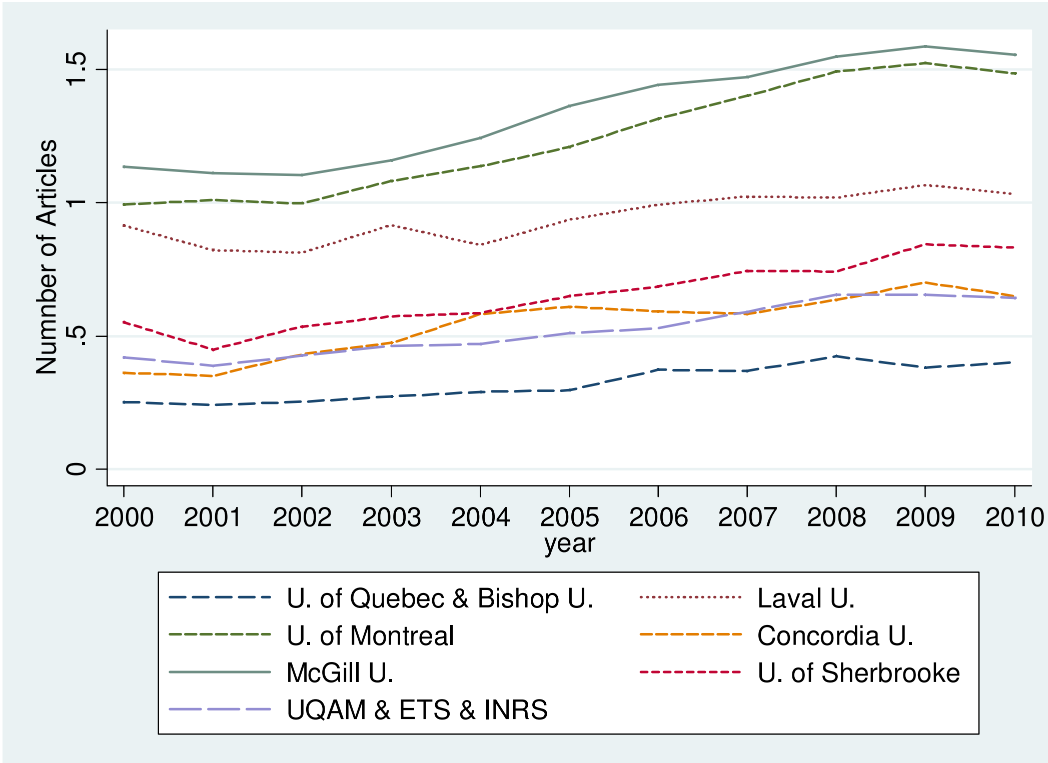

Figure 5.1: Trend of scientists' number of articles in universities

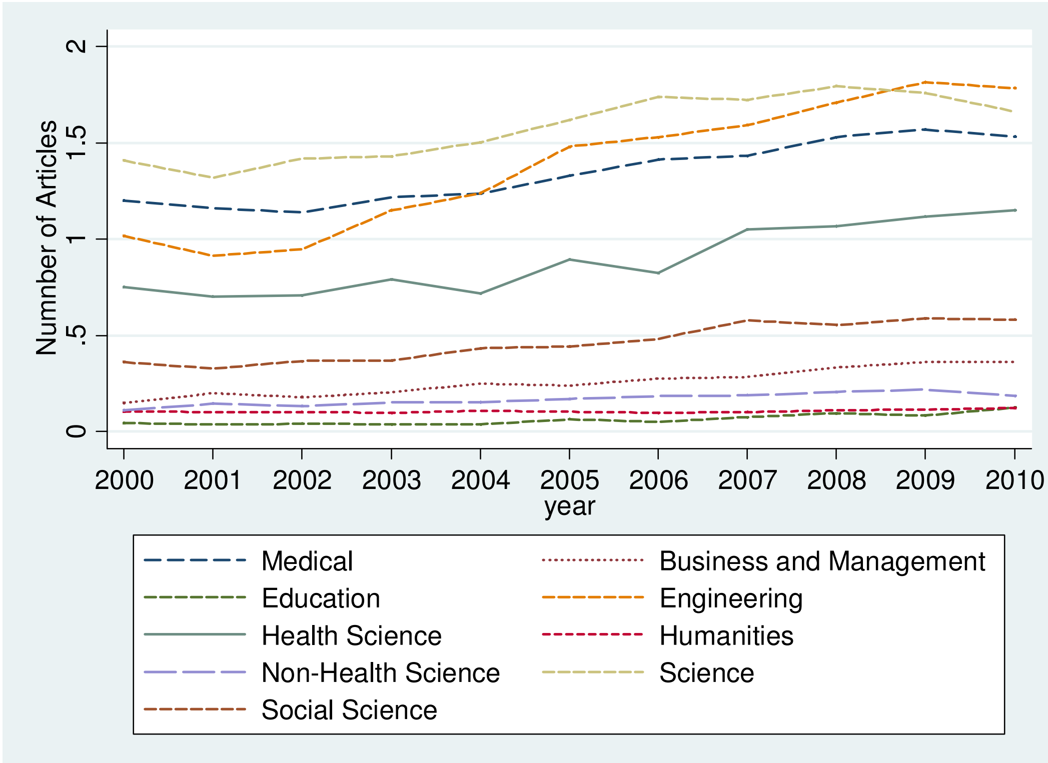

Figure 5.2: Trend of scientists' number of articles in different divisions

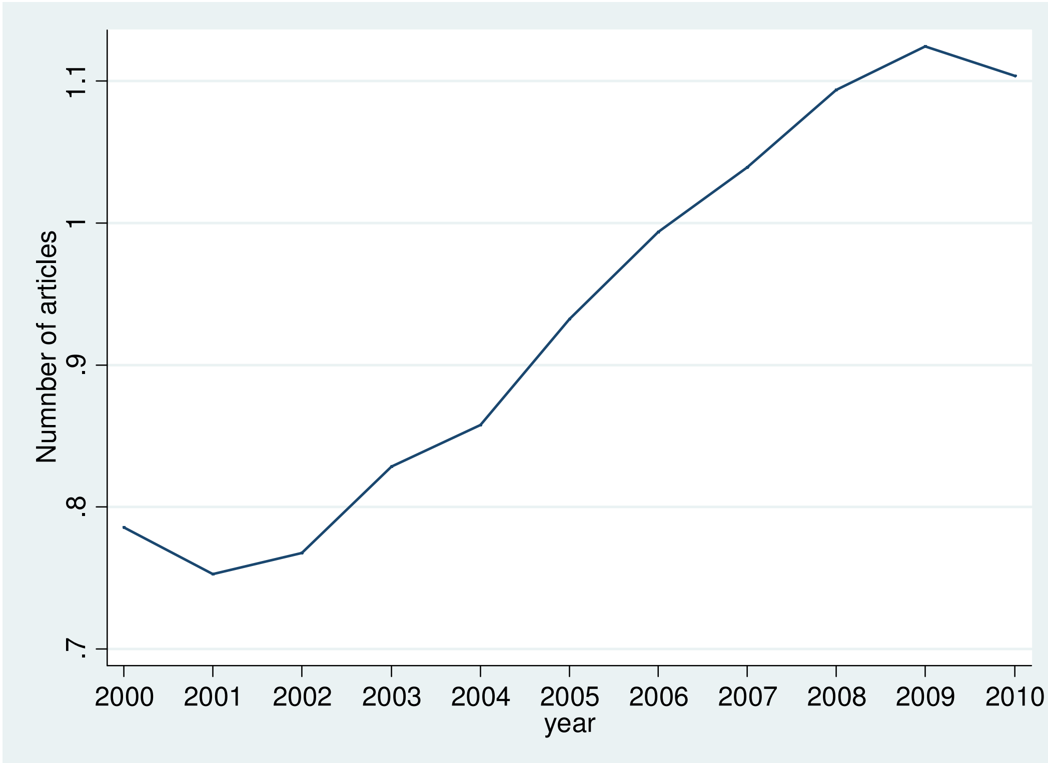

Figure 5.3: Average number of articles in each year

The possible reason behind yearly differences is that science and research policies may change over time. Hence funding and institutional support may have different formats over the years. In addition, the norms and standards for publication and motivations for research are different across divisions and universities. University dummies and research division dummies may thus account for part of the impact on scientific productivity[33]. Considering the mentioned explanatory variables, the resulting model is given by[34]:

$$\ln\left(nbArticle_{it}+1\right)=f\left(\begin{array}{l}\ln\left(PublicFundingO_{it}+1\right),\ln\left(PrivateFundingO_{it}+1\right),\\ \ln\left(NFPFundingO_{it}+1\right),\ln\left(nbAuthor_{it}\right),\\ (dIndChair\mid GCChair\mid dCRC\mid dIndGCChair\mid dChair)_{it},\\ dFemale_{i},Age_{it},Age_{it}^{\;2},D_{Field},D_{University},D_{Year}\end{array}\right)$$

---

[33] The "Year 2000", "McGill University", and the research division of "Medical science" are selected as reference points and are thus the excluded dummy variables.

[34] It should be noted that variables measuring funding and number of articles are transformed by natural logarithm function to have normal distribution and satisfy the necessary conditions for running regression equations.

It is important to note that two variables [*PublicfundingO*] and [*nbArticle*] are determined by each other and co-evolve over time, which is a potential source of endogeneity. As a consequence, ordinary least square regressions (for panel data) are biased. The main reason for this potential endogeneity is that scientists are assessed for public funding based on their CV and past productivity while at the same time, publication and research impact significantly depends on the funding available to researchers. We have also conducted Durbin–Wu–Hausman test for endogeneity, in which the residuals of the endogenous right-hand side variable (as a function of all exogenous variables) are put in the original panel OLS model. By showing that the coefficient of the residuals is significant, it is possible to conclude about existence of endogeneity[35].

Using instrumental variables (IV) is a common suggested method in literature to address endogeneity problems. There are two main requirements for using IVs: (1) the instruments must be correlated with the endogenous variable, and (2) the instruments should not be correlated with the error term in the main regression equation, which means that the instruments must not suffer from the same endogeneity problem. If there is more than one instrument for an endogenous variable, it is necessary to perform a two-stage regression, in which the first stage estimates the endogenous variable (the instrumented variable) based on a list of instrumental variables and the independent variables of the second-stage regression. Such an estimation removes the error term of the first stage and keeps the estimated amount for the second stage. Neglecting the error term of the endogenous variable and putting the estimated amount in the main regression equation should generally solve the endogeneity problem.

A number of instruments for the amount of public funding were tested during the course of this research and three were retained. The number of scientists in a university [*nbScientistUni*] is an indication of university size. It is expected that a university with a higher number of scientists may benefit from sharing of research costs and expenditures, which reduces the need for larger amounts of individual funding, hence benefitting from some kind of economies of scale.

The rank of previous funding obtained by an individual scientist can predict the future amount of funding. We would call this an echo effect, indicating that highly funded scientists are better able to find new sources of research funds in the future. The rationale behind this argument is that

[35] The test is reported in the note of each regression table in section 3.

successful researchers have an effective networking capacity (Winter et al., 2006) to generate and benefit from funding opportunities in a country such as Canada (Salazar and Holbrook, 2007). The difference between funding and rank of funding is that the rank of funding is just an ordinal variable and as such does not have information about amounts of funding (it meets the requirements of being an instruments, benefit from the panel format of the data without imposing an autoregressive structure to the model), while the amount of funding is informative about capacity of conducting original research. Therefore, we can use the rank of funding as an instrument of funding amount. The rank of a scientist in his/her research division in terms of three-year average amount of funding for the purpose of operational costs and direct expenditure of research [*PubORank*] is the second instrument retained in this research. Using the same rationale we can use the rank of a scientist in his/her research division in terms of three-year average number of articles [*PublRank*] as the third instrument. Regarding the mentioned instruments, in the first stage, the amount of public funding [*PublicfundingO*] is estimated by the instruments with one-year lag to avoid simultaneity problems. Considering the mentioned explanatory variables, the resulting model is given by:

$$1^{st}\ \text{Stage: } \ln\left(PublicFundingO_{it}+1\right)=g\begin{pmatrix} PubORank_{it-1}, PublRank_{it\text{-}1}, \ln\left(nbScientistUni_{it-1}\right), \\ variables\ \ from\ second\ stage \end{pmatrix}$$

$$2^{nd}\ \text{Stage:} \ln\left(nbArticle_{it}+1\right)=f\begin{pmatrix} \ln\left(PublicFundingO_{it}+1\right), \ln\left(PrivateFundingO_{it}+1\right), \\ \ln\left(NFPFundingO_{it}+1\right), \ln\left(nbAuthor_{it}\right), \\ (dIndChair \mid GCChair \mid dCRC \mid dIndGCChair \mid dChair)_{it}, \\ dFemale_i, Age_{it}, Age_{it}^{\ 2}, D_{Field}, D_{University}, D_{Year} \end{pmatrix}$$

In a well-specified model, the RHS variables (including instrumental variables) should not be highly correlated with each other. The low correlation refers to good level of independence and explanatory power of RHS variables. The correlation matrix is reported in the appendices section and the correlation coefficients are acceptable for estimating the regression equations and are thus appropriate choices from this point of view.

The main purpose of this research is to show whether holding a research chair as an external support is important and significant in promoting scientific publication. To test the first hypothesis, it is sufficient to run the two-stage panel regression on the entire data set to determine whether ‘holding a research chair’ is a significant RHS variable, either as a real cause or a channel for other variables/causes.

According to the chair characteristics, the networking and prestige effect of 'holding a research chair' may be mixed with the effect of funding. To address this issue, we use a matching technique and compare two 'chaired' and 'non-chaired' scientists who have similar funding, the same gender, and work in the same research field as each other. Following the methodology employed by Bérubé and Mohnen (2009), it is possible to find pairs of chair-holders and non-chair-holders by using the psmatch2 command in Stata and then to remove the unmatched records. The selection is made by generating propensity scores and choosing the pairs of scientists with the closest scores to each other[36]. The new data set consists of 'twin' scientists who are similar to each other in terms of funding, gender, and research fields.

Controlling with funding, gender, and research field, and keeping only the matched scientists in the regressions allows to disentangle the prestige effect from the funding effect of the chair and hence, 'holding a research chair' becomes a better and more informative signal for the prestige of scientists. In this case, the effect of 'holding a chair' on scientific productivity does not include funding effect or it is not related to the field or gender of the scientist. To test the second hypothesis, only matched pairs of scientists are being used in the regression analysis to identify whether holding a research chair has a significant effect on scientific productivity.

One of the important stages of the matching technique consists in validating the quality of matching. This implies that there should be no difference between the averages of the selection criteria (gender, funding, and research fields) when the comparison is made between chair holders and non-chair holders among the matched pairs. There can however be a difference when the comparison is made between the original database and the matched database. Table 5.1 summarizes these comparisons and shows that the matching is of acceptable quality for *dCRC*, *dIndGCChair*, and *dchair*.

[36] According to Stata handbook, "psmatch2 implements full Mahalanobis matching and a variety of propensity score matching methods to adjust for pre-treatment observable differences between a group of treated and a group of untreated. By default psmatch2 calculates approximate standard errors on the treatment effects assuming independent observations, fixed weights, homoskedasticity of the outcome variable within the treated and within the control groups and that the variance of the outcome does not depend on the propensity score. It is very important to note that the sort order of your could affect the results when using nearest-neighbor matching on a propensity score estimated with categorical (non-continuous) variables. Or more in general when there are untreated with identical propensity scores" (http://repec.org/bocode/p/psmatch2.html).

Table 5.1: Mean comparison between holders and non-holders of chairs for the matched and non-matched samples to show the quality of the matching

	Gender	Funding	Field[37]		Nb. of scientists
			NSE	Health	
dCRC=1	0.2000	$463,465	0.2779	0.1661	293
dCRC=0 (before matching)	0.2963	$85,994	0.2516	0.2262	7356
Significance level of difference[38]	***	***	>0.1	***	
dCRC=0 (after matching)	0.1023	$403,051	0.3583	0.1808	293
Significance level of difference	***	>0.1	>0.1	>0.1	
dIndGCChair=1	0.1319	$351,835	0.743	0.0694	144
dIndGCChair=0 (before matching)	0.2957	$95,728	0.2432	0.2268	7508
Significance level of difference	***	***	***	***	
dIndGCChair=0 (after matching)	0.1111	$369,080	0.6944	0.0902	144
Significance level of difference	>0.1	>0.1	>0.1	>0.1	
dchair=1	0.1809	$420,693	0.4238	0.1357	418
dchair=0 (before matching)	0.2991	$81,953	0.2427	0.229	7231
Significance level of difference	***	***	***	***	
dchair=0 (after matching)	0.1483	$364,117	0.4880	0.1770	418
Significance level of difference	>0.1	>0.1	>0.1	>0.1	

5.5 Result and discussion

Based on the models presented in methodology section, we first need to estimate the regressions on the entire dataset (tables 5.2, 5.3, and 5.4) to show that all types of chair have a positive and significant effect on scientific productivity. Then, after keeping only the matched scientists in the dataset, who are similar to each other in terms of gender, funding, and research field, the regression results indicate a significant and positive result only for the Canada research chair (tables 5.5, 5.6, and 5.7). Industrial chairs and chairs appointed by Canada research council (NSERC, SSHRC, and CIHR) do not have an independent positive effect on scientific productivity. However holding any kind of research chair [*dChair*] still has significant and positive effect mainly due to higher number of Canada research chairs included in the *dChair* dummy variable.

[37] The matching is done based on the dummy variables of NSE (Natural Science and Engineering) and HEALTH (Medical and Health Science). The third dummy is considered as the reference point.

[38] To find out whether two variables are indifferent, we validated three null hypotheses: inequality, greater amount, smaller amount. If any of the hypotheses is validated at 1%, then we can say that the difference is significant at the level of 1% (***). Otherwise, the difference is not significant.

Table 5.2: The second stage regression results over the entire sample (*dCRC*)

	xtivreg1	***xtivreg2***	***xtivreg3***	***xtivreg4***	***xtivreg5***	***xtreg6***
ln(PublicfundingO$_{it}$)	0.0227 ***	0.0210 ***	0.0210 ***	0.0211 ***	0.0211 ***	0.0083 ***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0005)
ln(PrivatefundingO$_{it}$)	0.0085 ***	0.0084 ***	0.0084 ***	0.0077 ***	0.0077 ***	0.0094 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
ln(NFPfundingO$_{it}$)	0.0059 ***	0.0058 ***	0.0058 ***	0.0058 ***	0.0058 ***	0.0066 ***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
dFemale$_i$	-0.0591 ***	-0.0546 ***	-0.0499 ***	-0.0548 ***	-0.0506 ***	-0.0666 ***
	(0.0076)	(0.0075)	(0.0076)	(0.0075)	(0.0076)	(0.0071)
Age$_{it}$	-0.0054 **	-0.0038 *	-0.0039 *	-0.0038 *	-0.0039 *	0.0049 ***
	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0018)
Age$_{it}^2$	3.30E-05 *	2.00E-05	2.10E-05	2.10E-05	2.10E-05	-1.10E-04 ***
	(2.0E-05)	(2.0E-05)	(2.0E-05)	(2.0E-05)	(2.0E-05)	(1.7E-05)
ln(nbAuthor$_{it}$)	0.4656 ***	0.4654 ***	0.4654 ***	0.4657 ***	0.4657 ***	0.4855 ***
	(0.0036)	(0.0036)	(0.0036)	(0.0036)	(0.0036)	(0.0032)
dCRC$_{it}$		0.2449 ***	0.2760 ***	0.2021 ***	0.2321 ***	0.2414 ***
		(0.0173)	(0.0192)	(0.0192)	(0.0213)	(0.0198)
dCRC$_{it}$*dFemale$_i$			-0.1529 ***		-0.1358 ***	-0.1311 ***
			(0.0420)		(0.0420)	(0.0393)
dCRC$_{it}$*ln(PrivatefundingO$_{it}$)				0.0114 ***	0.0108 ***	0.0095 ***
				(0.0023)	(0.0023)	(0.0021)
Constant	0.0236	-0.0166	-0.0152	-0.0156	-0.0143	0.0406
	(0.0557)	(0.0554)	(0.0554)	(0.0553)	(0.0553)	(0.0483)
Number of observations	80775	80775	80775	80775	80775	88422
Number of scientists	7651	7651	7651	7651	7651	7660
χ^2	40352.2 ***	41196 ***	41247 ***	41369.1 ***	41398.8 ***	47253.6 ***
R^2 within groups	0.1854	0.1868	0.1868	0.1866	0.1867	0.2079
R^2 overall	0.5456	0.5503	0.5505	0.5509	0.5511	0.5512
R^2 between groups	0.7631	0.7681	0.7684	0.7692	0.7695	0.7562

*, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 10.55, and 12 respectively. The amount of χ^2 for Durbin–Wu–Hausman test is 4901.85 and significant at level of 0.01, implying that endogeneity exists.

Table 5.3: The second stage regression results over the entire sample (*dIndGCChair*)

	xtivreg1	*xtivreg2*	*xtivreg3*	*xtivreg4*	*xtivreg5*	*xtreg6*
ln(PublicfundingO$_{it}$)	0.0227 ***	0.0224 ***	0.0224 ***	0.0224 ***	0.0224 ***	0.0087 ***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0005)
ln(PrivatefundingO$_{it}$)	0.0085 ***	0.0083 ***	0.0083 ***	0.0082 ***	0.0082 ***	0.0097 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0005)
ln(NFPfundingO$_{it}$)	0.0059 ***	0.0058 ***	0.0058 ***	0.0059 ***	0.0059 ***	0.0067 ***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
dFemale$_i$	-0.0591 ***	-0.0591 ***	-0.0580 ***	-0.0591 ***	-0.0580 ***	-0.0736 ***
	(0.0076)	(0.0076)	(0.0077)	(0.0076)	(0.0077)	(0.0071)
Age$_{it}$	-0.0054 **	-0.0054 **	-0.0054 **	-0.0054 ***	-0.0054 **	0.0038 **
	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0018)
Age$_{it}^2$	3.30E-05 *	3.30E-05 *	3.20E-05	3.30E-05 *	3.30E-05 *	-1.04E-04 ***
	(2.0E-05)	(2.0E-05)	(2.0E-05)	(2.0E-05)	(2.0E-05)	(1.7E-05)
ln(nbAuthor$_{it}$)	0.4656 ***	0.4658 ***	0.4658 ***	0.4658 ***	0.4658 ***	0.4870 ***
	(0.0036)	(0.0036)	(0.0036)	(0.0036)	(0.0036)	(0.0032)
dIndGCChair$_{it}$		0.1084 ***	0.1207 ***	0.0762 **	0.0886 ***	0.1157 ***
		(0.0247)	(0.0264)	(0.0330)	(0.0343)	(0.0318)
dIndGCChair$_{it}$*dFemale$_i$			-0.0935		-0.0926	-0.1187 *
			(0.0704)		(0.0704)	(0.0657)
dIndGCChair$_{it}$*ln(PrivatefundingO$_{it}$)				0.0043	0.0043	0.0048 *
				(0.0029)	(0.0029)	(0.0027)
Constant	0.0236	0.0281	0.0275	0.0294	0.0288	0.0810
	(0.0557)	(0.0557)	(0.0557)	(0.0557)	(0.0557)	(0.0485)
Number of observations	80775	80775	80775	80775	80775	88422
Number of scientists	7651	7651	7651	7651	7651	7660
χ^2	40352.2 ***	40397 ***	40403.2 ***	40426 ***	40431.8 ***	46374 ***
R^2 within groups	0.1854	0.1856	0.1856	0.1857	0.1857	0.2083
R^2 overall	0.5456	0.5460	0.5460	0.5461	0.5461	0.5464
R^2 between groups	0.7631	0.7633	0.7634	0.7634	0.7635	0.7491

*, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 10.55, and 12 respectively. The amount of χ^2 for Durbin–Wu–Hausman test is 5114.92 and significant at level of 0.01, implying that endogeneity exists.

Table 5.4: The second stage regression results over the entire sample (*dChair*)

	xtivreg1	*xtivreg2*	*xtivreg3*	*xtivreg4*	*xtivreg5*	*xtreg6*
ln(PublicfundingO$_{it}$)	0.0227 ***	0.0208 ***	0.0208 ***	0.0208 ***	0.0208 ***	0.0082 ***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0005)
ln(PrivatefundingO$_{it}$)	0.0085 ***	0.0080 ***	0.0080 ***	0.0073 ***	0.0073 ***	0.0089 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
ln(NFPfundingO$_{it}$)	0.0059 ***	0.0057 ***	0.0057 ***	0.0057 ***	0.0057 ***	0.0065 ***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
dFemale$_i$	-0.0591 ***	-0.0554 ***	-0.0505 ***	-0.0557 ***	-0.0512 ***	-0.0669 ***
	(0.0076)	(0.0075)	(0.0077)	(0.0075)	(0.0077)	(0.0072)
Age$_{it}$	-0.0054 **	-0.0040 *	-0.0041 **	-0.0041 **	-0.0042 **	0.0046 **
	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0018)
Age$_{it}^2$	3.30E-05 *	2.10E-05	2.20E-05	2.20E-05	2.20E-05	-1.09E-04 ***
	(2.0E-05)	(2.0E-05)	(2.0E-05)	(2.0E-05)	(2.0E-05)	(1.7E-05)
ln(nbAuthor$_{it}$)	0.4656 ***	0.4657 ***	0.4657 ***	0.4660 ***	0.4659 ***	0.4856 ***
	(0.0036)	(0.0036)	(0.0036)	(0.0036)	(0.0036)	(0.0032)
dChair$_{it}$		0.2091 ***	0.2315 ***	0.1761 ***	0.1985 ***	0.2112 ***
		(0.0147)	(0.0162)	(0.0170)	(0.0185)	(0.0172)
dChair$_{it}$*dFemale$_i$			-0.1207 ***		-0.1099 ***	-0.1132 ***
			(0.0367)		(0.0368)	(0.0344)
dChair$_{it}$*ln(PrivatefundingO$_{it}$)				0.0071 ***	0.0066 ***	0.0066 ***
				(0.0018)	(0.0018)	(0.0017)
Constant	0.0236	-0.0024	-0.0015	0.0025	0.0031	0.0551
	(0.0557)	(0.0554)	(0.0553)	(0.0553)	(0.0553)	(0.0483)
Number of observations	80775	80775	80775	80775	80775	88422
Number of scientists	7651	7651	7651	7651	7651	7660
χ^2	40352.2 ***	41140 ***	41180.3 ***	41243.4 ***	41269.6 ***	47152.8 ***
R^2 within groups	0.1854	0.1870	0.1870	0.1870	0.1870	0.2082
R^2 overall	0.5456	0.5501	0.5503	0.5505	0.5506	0.5508
R^2 between groups	0.7631	0.7677	0.7680	0.7684	0.7686	0.7554

*, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 10.55, and 12 respectively. The amount of χ^2 for Durbin–Wu–Hausman test is 4872.39 and significant at level of 0.01, implying that endogeneity exists.

Table 5.5: The second stage regression results over matched scientists (*dCRC*)

	xtivreg1	*xtivreg2*	*xtivreg3*	*xtivreg4*	*xtivreg5*	*xtreg6*
ln(PublicfundingO$_{it}$)	0.0278 ***	0.0254 ***	0.0254 ***	0.0255 ***	0.0255 ***	0.0096 ***
	(0.0074)	(0.0075)	(0.0075)	(0.0075)	(0.0075)	(0.0037)
ln(PrivatefundingO$_{it}$)	0.0032	0.0035 *	0.0035 *	-0.0001	0.0000	0.0030
	(0.0021)	(0.0021)	(0.0021)	(0.0027)	(0.0027)	(0.0025)
ln(NFPfundingO$_{it}$)	0.0014	0.0016	0.0016	0.0016	0.0016	0.0016
	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0018)
dFemale$_i$	-0.0906 *	-0.0994 *	-0.0480	-0.1001 **	-0.0584	-0.1266 *
	(0.0514)	(0.0509)	(0.0759)	(0.0509)	(0.0761)	(0.0689)
Age$_{it}$	0.0413 ***	0.0441 ***	0.0437 ***	0.0441 ***	0.0438 ***	0.0500 ***
	(0.0112)	(0.0112)	(0.0112)	(0.0112)	(0.0112)	(0.0096)
Age$_{it}^2$	-4.45E-04 ***	-4.63E-04 ***	-4.59E-04 ***	-4.64E-04 ***	-4.61E-04 ***	-6.60E-04 ***
	(1.1E-04)	(1.1E-04)	(1.1E-04)	(1.1E-04)	(1.1E-04)	(9.4E-05)
ln(nbAuthor$_{it}$)	0.6047 ***	0.6048 ***	0.6047 ***	0.6044 ***	0.6043 ***	0.5991 ***
	(0.0184)	(0.0183)	(0.0183)	(0.0183)	(0.0183)	(0.0163)
dCRC$_{it}$		0.1297 ***	0.1445 ***	0.0939 **	0.1068 **	0.0739 *
		(0.0396)	(0.0427)	(0.0430)	(0.0464)	(0.0417)
dCRC$_{it}$*dFemale$_i$			-0.0895		-0.0726	-0.0542
			(0.0978)		(0.0982)	(0.0892)
dCRC$_{it}$*ln(PrivatefundingO$_{it}$)				0.0084 **	0.0082 **	0.0074 **
				(0.0039)	(0.0039)	(0.0036)
Constant	-1.2490 ***	-1.3562 ***	-1.3553 ***	-1.3358 ***	-1.3356 ***	-0.7559 ***
	(0.2877)	(0.2881)	(0.2881)	(0.2882	(0.2882	(0.2448)
Number of observations	6393	6393	6393	6393	6393	6979
Number of scientists	586	586	586	586	586	586
χ^2	3159.02 ***	3200.45 ***	3201.24 ***	3205.26 ***	3205.25 ***	3416.71 ***
R^2 within groups	0.2527	0.2533	0.2532	0.2536	0.2535	0.2521
R^2 overall	0.5332	0.5384	0.5387	0.5392	0.5394	0.5093
R^2 between groups	0.6986	0.7055	0.7060	0.7068	0.7071	0.6484

*, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 2, 10.9, and 12 respectively. The amount of χ^2 for Durbin–Wu–Hausman test is 323.5 and significant at level of 0.01, implying that endogeneity exists.

Table 5.6: The second stage regression results over matched scientists (*dIndGCChair*)

	xtivreg1	*xtivreg2*	*xtivreg3*	*xtivreg4*	*xtivreg5*	*xtreg6*
ln(PublicfundingO$_{it}$)	0.0384 ***	0.0383 ***	0.0381 ***	0.0393 ***	0.0392 ***	0.0171 ***
	(0.0093)	(0.0094)	(0.0094)	(0.0095)	(0.0095)	(0.0049)
ln(PrivatefundingO$_{it}$)	0.0042	0.0043	0.0043	0.0097 **	0.0098 **	0.0103 ***
	(0.0032)	(0.0032)	(0.0032)	(0.0043)	(0.0043)	(0.0040)
ln(NFPfundingO$_{it}$)	0.0018	0.0018	0.0018	0.0017	0.0017	0.0023
	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0025)
dFemale$_i$	-0.1661 **	-0.1650 *	-0.1462	-0.1623 *	-0.1373	-0.2138 *
	(0.0841)	(0.0843)	(0.1275)	(0.0845)	(0.1279)	(0.1136)
Age$_{it}$	-0.0045	-0.0044	-0.0041	-0.0039	-0.0036	0.0172
	(0.0176)	(0.0176)	(0.0176)	(0.0176)	(0.0176)	(0.0144)
Age$_{it}^2$	-4.20E-05	-4.30E-05	-4.50E-05	-4.70E-05	-5.00E-05	-3.51E-04 ***
	(1.7E-04)	(1.7E-04)	(1.7E-04)	(1.7E-04)	(1.7E-04)	(1.4E-04)
ln(nbAuthor$_{it}$)	0.5046 ***	0.5041 ***	0.5043 ***	0.5062 ***	0.5065 ***	0.5284 ***
	(0.0261)	(0.0262)	(0.0262)	(0.0262)	(0.0262)	(0.0228)
dIndGCChair$_{it}$		-0.0131	-0.0097	0.0496	0.0543	0.0470
		(0.0587)	(0.0610)	(0.0691)	(0.0714)	(0.0630)
dIndGCChair$_{it}$*dFemale$_i$			-0.0333		-0.0441	-0.0365
			(0.1677)		(0.1681)	(0.1497)
dIndGCChair$_{it}$*ln(PrivatefundingO$_{it}$)				-0.0104 **	-0.0104 **	-0.0050
				(0.0059)	(0.0059)	(0.0053)
Constant	0.0140	0.0134	0.0072	-0.0436	-0.0516	0.0580
	(0.4565)	(0.4566)	(0.4578)	(0.4568)	(0.4580)	(0.3821)
Number of observations	3234	3234	3234	3234	3234	3522
Number of scientists	288	288	288	288	288	288
χ^2	1253.64 ***	1253.36 ***	1252.08 ***	1253.15 ***	1251.95 ***	1391.07 ***
R^2 within groups	0.2256	0.2256	0.2257	0.2261	0.2261	0.2401
R^2 overall	0.4303	0.4305	0.4306	0.4302	0.4303	0.4079
R^2 between groups	0.6001	0.6002	0.6001	0.5992	0.5992	0.5381

*, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 3, 11.22, and 12 respectively. The amount of χ^2 for Durbin–Wu–Hausman test is 135.45 and significant at level of 0.01, implying that endogeneity exists.

Table 5.7: The second stage regression results over matched scientists (*dChair*)

	xtivreg1	*xtivreg2*	*xtivreg3*	*xtivreg4*	*xtivreg5*	*xtreg6*
$ln(PublicfundingO_{it})$	0.0388 ***	0.0367 ***	0.0368 ***	0.0366 ***	0.0367 ***	0.0119 ***
	(0.0054)	(0.0054)	(0.0054)	(0.0054)	(0.0054)	(0.0028)
$ln(PrivatefundingO_{it})$	0.0068 ***	0.0068 ***	0.0068 ***	0.0097 ***	0.0097 ***	0.0126 ***
	(0.0018)	(0.0018)	(0.0018)	(0.0024)	(0.0024)	(0.0022)
$ln(NFPfundingO_{it})$	0.0016	0.0017	0.0017	0.0017	0.0017	0.0029
	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0015)
$dFemale_i$	-0.1016 **	-0.1076 **	-0.0794	-0.1075 **	-0.0740	-0.1441 ***
	(0.0422)	(0.0419)	(0.0615)	(0.0419)	(0.0616)	(0.0554)
Age_{it}	0.0187 **	0.0204 **	0.0202 **	0.0209 **	0.0206 **	0.0380 ***
	(0.0092)	(0.0092)	(0.0092)	(0.0092)	(0.0092)	(0.0079)
Age_{it}^2	-2.24E-04 **	-2.40E-04 ***	-2.38E-04 ***	-2.43E-04 ***	-2.41E-04 ***	-5.24E-04 ***
	(8.9E-05)	(8.9E-05)	(8.9E-05)	(8.9E-05)	(8.9E-05)	(7.6E-05)
$ln(nbAuthor_{it})$	0.5553 ***	0.5569 ***	0.5568 ***	0.5576 ***	0.5575 ***	0.5669 ***
	(0.0149)	(0.0149)	(0.0149)	(0.0149)	(0.0149)	(0.0131)
$dChair_{it}$		0.1195 ***	0.1273 ***	0.1462 ***	0.1562 ***	0.1540 ***
		(0.0321)	(0.0344)	(0.0353)	(0.0378)	(0.0342)
$dChair_{it}*dFemale_i$			-0.0515		-0.0612	-0.0425
			(0.0820)		(0.0823)	(0.0743)
$dChair_{it}*ln(PrivatefundingO_{it})$				-0.0058 *	-0.0059 *	-0.0043
				(0.0032)	(0.0032)	(0.0030)
Constant	-0.7362 ***	-0.7928 ***	-0.7921 ***	-0.8208 ***	-0.8208 ***	-0.5576 ***
	(0.2365)	(0.2361)	(0.2361)	(0.2366)	(0.2366)	(0.2015)
Number of observations	9092	9092	9092	9092	9092	9928
Number of scientists	836	836	836	836	836	836
χ^2	4290.62 ***	4334.55 ***	4333.85 ***	4336.14 ***	4335.52 ***	4721.13 ***
R^2 within groups	0.2426	0.2437	0.2436	0.2441	0.2441	0.2516
R^2 overall	0.5054	0.5097	0.5098	0.5095	0.5097	0.4877
R^2 between groups	0.6817	0.6872	0.6874	0.6865	0.6867	0.6388

*, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 2, 10.87, and 12 respectively. The amount of χ^2 for Durbin–Wu–Hausman test is 411.79 and significant at level of 0.01, implying that endogeneity exists.

The Durbin–Wu–Hausman test for endogeneity is also done for each set of regressions separately. As explained before, the test first generates the residuals of "public funding" (as a function of all exogenous variables) and then put them in the original panel OLS model[39]. By showing that the coefficient of the residuals is significant, the test revealed the existence of endogeneity. The test is reported in the note of each table, indicating that coefficients of the residuals are significant.

Our results clearly validate our first hypothesis but only partially validate our second hypothesis. One may question whether research chairs in general are an independent contributor to research productivity or whether they are simply a proxy for other known factors in literature. Considering

[39] For this test we use all variables that are available for regression, like variables in regression 5 of each table.

the literature and mentioned mission or mandate of research chairs, it is possible to argue that holding a chair increases the amount of funding available to scientists.

In our second hypothesis, we make a distinction between the effect of funding and holding a research chair. By matching pairs of chair holders and non-chair-holders who are similar to each other in terms of funding, gender, and research field, we can investigate other aspects of the impact of holding chair such as collaborating with brilliant talents. By estimating the regression model only on matched pairs of scientists, holding a chair cannot be a proxy for the matching criteria (funding, gender, and research field) anymore. Considering tables 5.5, 5.6, and 5.7, we can verify hypothesis 2 for industrial chairs and research chairs appointed by federal granting councils but this hypothesis cannot be validated for the 'Canada research chairs' because its effect is still positive and significant even after matching. In other words, same gender holders of Canada research chairs with equivalent funding and in the same research field as other scientists still generate a greater number of articles than these other scientists.

A number of factors can contribute to explaining this finding. The first is that the Canada research chair intends to be a prestigious research sign in Canada. Based on its mandate, the Canada research chair program aims to attract and retain some of most accomplished and promising minds in the world and it is awarded to scientists from all disciplines including engineering and the natural sciences, health sciences, humanities, and social sciences. It is more prestigious than any other research chairs, with the exception perhaps of the newly introduced Canada research excellence chairs, and the holders are expected to be more capable in expanding their collaborative research. Other scientists may also have more willingness to conduct collaborative research with the Canada research chair holders. In addition, the Canada research chair programme grants more visibility to the chair-holders and they can recruit more talented students and researchers.

The second explanation is that industrial chairs are appointed by firms to promote research and its application, probably with major benefits to the firms themselves and as such, serve an entirely different purpose. In other words, this type of chair is not necessarily and originally designed for the sake of scientific publication. The chairs appointed by research councils may have quite similar characteristics. Looking at these chairs' description, most of chair holders are appointed as industrial chair, partly funded by industry and by the relevant granting council. There is some evidence in the literature indicating that industrial funding forces researchers to shift to more applied research, neglecting their normative responsibilities for knowledge development (Geuna and Nesta, 2003; Partha and David, 1994).

In addition to the effect of holding a chair on scientific productivity, there are also some interesting results regarding the control variables in econometric model. Funding from different sources is always a positive and significant determinant of scientific productivity before matching[40]. After matching, however, the positive effect of funding is only significant for public funding and private funding. We also consider the effect of interaction between “holding a chair” and the amount of funding. From a technical point of view, it is not possible to estimate the interactive effect with an endogenous variable in 2SLS models because its amount is estimated in the first stage and we are not using the raw value reported in dataset. However, we can estimate the effect of the interaction of holding a chair with private funding[41].

The effect of this interactive variable is positive and significant for *dCRC* before and after matching. This suggests that private funding contributes to helping Canada research chair-holder increase their scientific productivity. In contrast, the funding and chair interaction is not significant for *dIndGCChair* before matching but negative and significant after matching, hence implying that both chair-holders and non-chair-holders positively benefit from private funding before matching but after matching this effect disappears. The same story is valid for *dChair*. The interactive effect of *dChair* and private funding is positive before matching but negative after matching. Looking precisely at the numbers, the results reveal that private funding has a positive effect before matching but its positive effect is be weakened or even disappears after matching (the combined effect of private funding with its interactive effect is shown in figure 5.4 for *dIndGCChair* and *dChair* after matching).

[40] Funding from the private sector and funding from the not-for-profit sector are directly put in the regression equation while funding from the public sector is first estimated by the instrumental variables and then inserted into the regression model. The first stage model regressions, reported in the appendices section, show the significant role of instrumental variables.

[41] The interaction with between the chair dummy variables and not-for-profit funding was tested and was never significant.

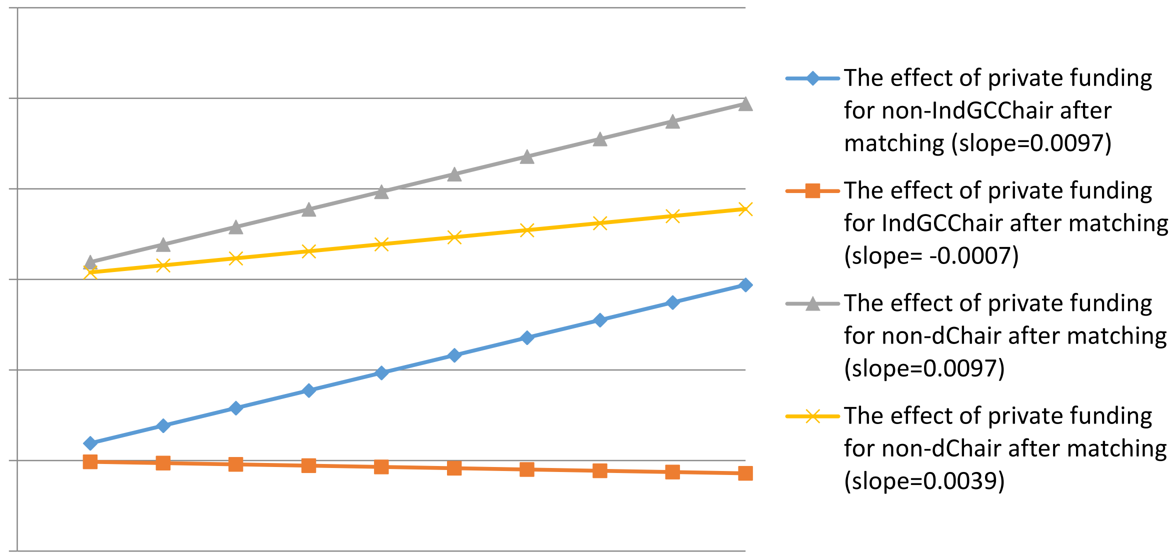

Figure 5.4: Comparison of private funding effect on scientific productivity for different values of *dIndGCChair* and *dChair*

Considering the fact that industrial chairs are partially funded by the private sector, we suspect that more private funding for industrial chair-holders may detract the chair holder from scientific traditional outputs and towards non-scientific and/or more technological and applied outputs such as patents. The negative effect of private funding on scientific productivity finds some echo in the literature (Goldfarb, 2008; Kleinman and Vallas, 2001).

The gender of a scientist has a negative and significant impact, implying that women are less likely to publish than men, another result vastly supported in the literature (Kyvik and Teigen, 1996; Leahey, 2006a; Long, 1990). In terms of interaction between chair and gender, we would argue that before matching, this interactive effect is negative and significant for *dCRC* and *dChair* but is non-significant after matching. This would tend to suggest that in general a female scientist may benefit less from holding a chair than her male colleagues. But when only matched pairs are kept in the dataset, the higher scientific productivity of male chair-holder compared to male non-chair-holder is not different from that of their female colleagues.

The age of the scientists seems to affect scientific productivity negatively before matching and with an inverted-U shape pattern (with a peak around 48 years old) after matching. Such life-cycle and investment motivated behaviour is acknowledged in the literature (Bonaccorsi and Daraio, 2003;

Diamond, 1986; Levin and Stephan, 1991). Age loses its significance when scientist pairs are matched based on *dIndGCChair*, possibly because of smaller size of data for this chair. The results for panel OLS regression are also provided to show that the results of our paper are robust enough (the sixth row of each table).

5.6 Conclusion

In this chapter we show that holding a research chair is a significant determinant of scientific publication when the regression is run over the entire data set of Quebec scientists. As previously explained, a distinction should be made to clarify different attributes of research chairs and their effect on scientific productivity. For instance, the research chair comes with its own funds so the question of interest is whether we can find a positive effect of research chairs on scientific productivity after controlling for the funding obtained by the chair holder. To investigate this relationship, we applied a matching technique to control for gender, funding and research field of chair and non-chair holders. This may indicate whether the effect of holding a research chair on scientific productivity is still significant after controlling for the mentioned attributes.

After such matching, the results show that the effect of the Canada research chair program on scientific productivity remains significant and positive while the effect of industrial chairs and the chairs appointed by Canada federal granting council (NSERC, and CIHR) become non-significant. This finding highlights the special attributes of the Canada research chair program, which are not replicated in other chairs. Those specific attributes may significantly push scientific productivity. Among others, Canada research chairs are generally associated with some degree of prestige or higher visibility to recruit talented students or to have research collaboration with top scientists in the field. The fact that other types of research chairs, once matched with equivalent scientists, do not have an impact on scientific output in terms of quantity, implies not that these chair holders are lesser scientists, but that they are devoting part of their time to other endeavours of a more practical nature. Hence universities are maintaining a balance between the pursuit of pure scientific knowledge and its application to socio-economic benefits. We are not going to shock anyone by stating that by solely studying scientific articles, we are missing a great deal of the role of university professors. Although not trivial, future research should aim to cast a wider net on outputs, outcomes and impacts of university research.

Our research has a number of limitations. First, the data has a number of missing entries, particularly regarding gender and age, there may be some selection bias introduced by this missing information. Second, our study only covers the province of Quebec and may not be generalised to other parts of the world. Third, in the 2SLS model, we used instruments for public funding, which were the best possible and accessible variables at the time of the study but there may be better instruments such as the size of the department or of the research group. Fourth, we used the number of articles as a measure of scientific productivity, which is appropriate but it cannot completely reflect the productivity of scientists. In addition, the Web of Science does not cover adequately the social sciences and humanities fields.

Let us now make a few suggestions for further studies. The first suggestion is to make a qualitative analysis on the attributes of the research chairs. Gathering data from interviews with chair-holders or surveys filled up by experts would shed some light on how holding a research chair contributes to scientific productivity. Such qualitative work could help investigate the effect of research chair from a social or psychological point of view. Moreover, it would also be possible to find the institutional effect of the research chairs or the impact of peripheral institutional settings on the productivity of research chairs.

The second suggestion is to make a comparison between the short-term and long-term effect of research chairs on scientific productivity. Although we understand that some types of research chairs may not have a significant effect on scientific publication (a short-term effect), our research did not make any investigation on how they form networks and accumulate research skills for future studies or how the chairs train the next generation of scientists.

Finally, a third suggestion would be to investigate other types of outputs for research chairs which are not devoted to purely academic endeavours but aimed at a more applied impact. The results for the impact of these types of chairs may be completely different. Our research is therefore limited by the variables used to measure output (of a purely scientific nature).

Chapter 6 GENERAL DISCUSSION

In this thesis, we investigated the main determinants of scientific production and research impact. Scientific production is measured by the number of articles and research impact is quantified by the citation count. To do so, we have integrated the data sets of researchers funding and publication in the province of Quebec. We have access to Thomson Reuters Web of Science database on scientific articles (2000-2012), which includes information about date of publication, journal name, authors, affiliations, and number of citation each article receives. Funding information of scientists comes from the Quebec University Research Information System (Système d'information sur la recherche universitaire or SIRU) of the Ministry of Education and Research (MELS). This database reports funding information including grants and contracts of all Quebec academics, on a yearly basis during the period 1985-2012. The age and gender of scientists was obtained from the MELS internal database.

In identifying the main determinants, we have used a two-stage regressions to address endogeneity issue in our papers. The amount of funding received and the number of articles published are co-evolved and determined by each other simultaneously. Such co-determination is a potential source of endogeneity, which biases the result of ordinary least square (OLS) regressions. The main explanation for such endogeneity is that the scientists are evaluated for public funding contest based on their CV and past effectiveness, while at the same time, publication and research impact depends on the funding capability of researchers.

Using instrumental variables (IV) is a common method in literature to deal with endogeneity issues. Instruments are required to be correlated with the endogenous variable, and should not be correlated with the error term in the main regression equation, which means that instruments should not induce another source of endogeneity. If there is more than one instrument, the endogenous variable is being estimated in a first-stage regression based on a list of instruments. Such estimation removes the error term (source of endogeneity) and keeps the estimated amount for the second stage.

In the first stage, the amount of public funding [*ln(PublicfundingO)*] is estimated by the following variables: (1) the number of articles is an important factor that measures the past productivity of scientists [*ln(nbArticleAvg3)*] as it is the main component of one's CV; (2) infrastructure related public funding is a proxy indicating how much a scientist is equipped to conduct research in the frontier of knowledge [*ln(PublicfundingO)*] ; (3) age and its square, which generally measures the a scientist' research experience [*Age*, Age^2].

As discussed in chapter 3, there is still a controversial debate in the literature about using citations as a measure research impact. Based on Aksnes (2006), citation counts correlate with own assessments of authors about their scientific contribution. However, the author notes that citation counts are not reliable at the level of the individual article and average citation rate of the discipline is a more robust measure than the journal's one. Citations are nonetheless a useful indicator for assessing the ranking of scientist (Radicchi et al., 2009).

In a different study, Kostoff (1998) argues that every citation comes from the combination of two main reasons: the real component of intellectual heritage and random components of self-interest. The author notes that the random effect exists but it disappears in the aggregation of citation counts and therefore the number of citations is a good measure for the "impact" of research. Phelan (1999) provides the same justification.

Some warning are however given by Amsterdamska and Leydesdorff (1989) about the use of citation count as a measure of research impact. The authors note that an article is being cited either to make a linkage with the current literature or to be used as an evidence for the line of reasoning in the paper. Therefore, a cited paper may be used in the citing papers with distinctive purposes and implications.

Moed et al. (1985) also argue that citations just refer to the impact and are not an indicator of research impact. Any publication should nonetheless have a minimal quality to generate a research impact. For instance, the visibility of journals and the extent to which researchers provide a public service are two factors that positively affect the citation but they are not a real and significant determinants of research impact. Furthermore, Olson et al. (2002) claim that even peer reviewed publications do not guarantee a minimum research impact because for example in medical fields, there is a bias for publishing research with positive results.

Assuming the number of citations as a measure for research impact, there are different determinants for the number of citations. First, our regressions show that scientific publications of female scientists are cited similarly as men's publication because the gender variable is not a significant determinant except for the research domain of humanities (it is mainly due to the small sample size of that domain). However, this contrasts with some evidence in the literature that points towards the relative under-performance of women in terms of number of articles and research impact.

A greater visibility of scientists, as determined by their number of articles or by the journal impact factor, is associated with a greater number of citations. In addition to the separate assessment of the effect of the number of articles or journal impact factor, our regression models have also tested the collaborative effect.

The regression shows that a higher number of articles in higher impact factor journals results in more citations than the same number of articles in less prestigious journals. The policy implication of our results should be to encourage scientists to publish in journals with a high impact factor, mainly because articles in theses journals are widely read and used by the scientific community. The journal impact factor is also a proxy for research impact because such journals have more paper submissions and editors have more options to choose higher impact papers.

The results also show that articles with more authors are generally more likely to be cited. The possible reasons behind this finding are related to the benefits of collective work, which increase the research impact, knowledge spillover or tacit knowledge transfer. The effect of the funding amount on the number of citations is domain dependent in a sense that in some domains it is insignificant and in some domains it is positive and significant. This non-significance does not necessarily imply that funding is ineffective for scientific productivity because the model just investigates the number of citations (not the number of articles). Considering the mixed effect of funding on the number of citations, which depends on the research domain, it is possible to propose domain-specific policies to improve the effectiveness of research funding on the number of citations.

In chapter 4, a contract theory model is developed to predict the effect of collaboration with star scientists. This model explains the cost-benefit process of a researcher in publishing a paper. We first reviewed distinctive definitions of star scientists in the literature. Both the effects of funding and collaboration on scientific productivity have been examined in literature. To the best of our knowledge, however, there is not a comprehensive study on the effect of collaborators' funding on scientific productivity. Based on the known correlation between funding and scientific productivity, we consider funding as a proxy for being star. We have introduced our definitions of star scientists as those in the top 25%, top 10%, or top 5% in terms of total funding, funding from the public sector, and funding from the private sector.

Having adequate funds is a necessary condition for researchers to buy instruments and hire research assistants in order to follow new ideas and conduct research at the frontier of knowledge. The results showed that collaboration with top funded scientists significantly contribute to the scientific productivity. Among different reasons, top funded scientists may have advanced and updated research tools and equipment that can improve scientific productivity. In addition, such collaboration may increase knowledge transfer, getting valuable experience, and learning new skills.

Economies of scale are the second possible reason justifying our results. The per-unit fixed cost of shared resources (for instance the cost of laboratory equipment) is lower in the case of research collaboration, which decreases the marginal cost of publication. The third explanation is related to the fact that researcher with more fund can have access to modern and up-to-date instrumentation. This leads to more options for conducting research. Neumann and Finaly-Neumann (1990) argued that research publication is directly influenced by support such as research instruments.

Chapter 5 exclusively investigated the effect of holding a research chair on scientific productivity. According to the chair features, the networking and prestige effect of ‘holding a research chair’ may be mixed with the effect of funding. It is possible to separate the two mentioned effect by comparing two ‘chaired’ and ‘non-chaired’ scientists who have similar funding, the same gender, and work in the same research field as each other (matching techniques). Matching chair and non-chair can be made by generating propensity scores and choosing the pairs of scientists with the closest scores to each other.

One of the important steps in matching technique is to control the quality of matching. This means that there should be no difference between the averages of the selection criteria (gender, funding, and research fields) when the comparison is made between chair holders and non-chair holders among the matched pairs. There can however be a difference when the comparison is made between the original database and the matched database. In chapter 5, we showed that the matching process has acceptable quality.

According to the matching technique, the regression models were first run on the entire dataset, which show that all types of chair have a positive and significant effect on scientific productivity. However, after matching when only matched scientists (who are similar to each other in terms of gender, funding, and research domain) are kept in the dataset, the regression results reveal a significant and positive result only for the Canada research chair while industrial chairs and chairs appointed by Canada research council (NSERC, SSHRC, and CIHR) become insignificant. We aimed to make a distinction between the effect of exclusive funding of chair and the effect of other attributes of holding a research chair. This aim is met when we compared two chair and non-chair scientists with the same amount of funding.

To investigate reason behind this observation, it is possible to argue that the Canada research chair is designed to be a prestigious research sign in Canada to attract and retain some of most accomplished and promising minds in the world. It stresses on the research expertise and capability. Moreover, it has more prestige than any other research chairs, which may lead to more visibility. Another difference between Canada research chair and other types of chair is that industrial chairs are appointed by firms to promote research and its application related to firm, which is not necessarily aimed for the sake of scientific publication while Canada research chairs are appointed by independent selection committee, in which conducting high quality research is one of the most important criteria.

Holding a research chair may also affect other determinants of scientific productivity. For instance, the positive effect of private funding is higher for Canada research chair-holders. There is also some sort of interaction between holding a research chair and the gender of a scientist. A female scientist may benefit less from holding a chair than her male colleagues. However, after matching and keeping only matched scientists in data set, this difference disappears.

In all of the three chapters, we introduced control variables to make the regression results more reliable and robust. These also deserve some remarks particularly regarding the effect of funding, gender, and age on scientific productivity and on research impact in our database. The effect of gender of a scientist is sometimes insignificant. But if it is significant, it has a negative impact, implying that women are less likely to publish than men, a result that is vastly supported in the literature (Kyvik and Teigen, 1996; Leahey, 2006a; Long, 1990). The age of a scientist also exhibits a significant inverted U-shaped effect. Scientists have increasing scientific productivity until a specific age but their productivity drops afterward. This is similar to the earlier works explaining the age effect (Bernier et al., 1975; Diamond, 1986).

Considering the size of research team, our results show scientists with more co-authors are expected to have more articles and better quality of research. This finding is compatible with evidence in literature indicating positive effect of collaboration (Johnes, 1988; Melin, 1996). To justify such finding, in a research team with numerous researchers, tasks are done collectively and cooperatively. It probably leads to knowledge spillover or tacit knowledge transfer, which increases the capacity and capability of researchers in conducting original research. Our model also verifies the effect of university and research discipline in addition to the year-specific effect on scientific productivity and research impact.

In terms of funding, it does not have a major effect on the number of citations. For example, public funding has a positive significant effect only for science and social science. The effect of private funding is significant only for the medical science. The non-for-profit funding effect is not significant at all. The empirical and theoretical evidence in literature supporting the positive effect of funding on productivity of scientists are known, but our results indicate that higher funding does not necessarily results in publications which are more cited. This does not contradict the positive effect of funding on scientific productivity but it indicates that higher funding is not a determinant of article citation counts.

In terms of scientific productivity, our results indicate the positive effect of funding on the number of articles. All types of funding from different sources (funding from the public, private, and not-for-profit sectors) are all positively and significantly increase the number of scientific publications: the more funded the scientists are the more articles they publish. As a simple justification, it is possible to argue that sufficient financial support is a necessary condition for researchers to buy instruments and hire assistants in order to follow new ideas and conduct research at the frontier of knowledge. Our findings hence echo past studies in terms of public funding (Feldman and Graddy-Reed, 2013; Salter and Martin, 2001) and private funding (Manjarrés-Henríquez et al., 2009; Manjarrés-Henríquez et al., 2008).

Chapter 7 **CONCLUSION AND RECOMMENDATIONS**

This thesis studied the determinants of number of citations, the effect of having a research collaboration with top-funded scientists on scientific productivity, and the effect of holding a research chair on scientific productivity. In chapter 3 we have concluded that the number of articles and the visibility of a researcher, the impact factor of the journal, the size of the research team, and the institutional setting of the university are the important determinants of citation counts. However, we have found that there is no significant effect of public funding and gender in most of the domains examined.

Assuming that the number of citations is a good measure for research impact and, in turn, for a certain kind of quality, we proposed some policy advice to address the issues discussed in the chapter. First, it seems that collaborative works (measured by the size of research team) can influence the impact of the research, and policy makers should therefore encourage research of a more collaborative nature. The measure of such research collaboration is not only limited to the number of authors in each articles but can also measure the quality, extent, and durability of collaboration. However, the only variable we have on hand, which measure the research collaboration, is the number of individuals in the authors list.

Second, the significance level of the funding effect on research impact (which is not the same for all domains) does not necessarily imply that funding is ineffective for the knowledge production process/chain – this paper only investigates researchers' scientific impact and not their research productivity, which his considered as an input here. There is strong evidence in the literature about the significant effect of different funding types on scientific productivity (Manjarrés-Henríquez et al., 2008; Pavitt, 2001; Salter and Martin, 2001) to which our research contributes: public funding has a positive and significant effect for "Science" and "Social Science", while the private funding effect is positive and significant for "Science" and "Medical Science". These results point towards domain-specific policies and incentives because the effect of funding is domain-dependent, *ceteris paribus.* Hence a domain-specific policy can be an effective tool for improving the research impact in specific domains, without the need for general policy making, which may require holistic manipulation of science policy.

The third policy implication is to incentivise researchers to publish in journals with a high impact factor. Such journals have more visibility and their articles are widely read and used by the scientific community, more so than articles in other journals with lower impact factors. Moreover, the impact factor of a journal is a proxy for research impact because journals with higher impact factor have more submissions and editors are able to choose higher quality papers. Although a greater number of articles in the past contributes to improving the visibility of articles in the future (and hence their perceived research impact), the positive and significant effect of the interactive variable, which measures the modulating effect of the number of articles on the journal impact factor, on research impact implies that past articles in journals with a higher impact factor can reflect the intrinsic research impact of individuals. We can also argue that there is a learning experience from the past collaboration, especially if that collaboration led to a highly cited paper.

The last but not the least, our research did not find any gender bias in terms of research impact (except for the domain of Humanities). This contrasts with some evidence in the literature that points towards the relative under-performance of women in terms of number of articles and research impact. However, it is not possible to make a policy conclusion in this regard because there may be a great number of reasons that may explain why women are less cited in one specific domain. First, they publish less and are thus less visible. Second, there may be conscious or unconscious discrimination. Third, women may be involved in more multidisciplinary research that is harder to publish. Fourth, women may spend more time involved in other duties at university. Although more investigations are needed, some incentive programs can be to encourage women to apply for more funding, and to go to more conferences (to be more visible). In addition, mentorship programs should be put in place where the gap is significant.

Chapter 4 developed a theoretical model and extract some hypotheses about the effect of collaboration with top-funded scientists on the scientific productivity. It then validated the hypotheses with empirical analysis and showed that such collaboration has a positive effect on scientific productivity. This significant effect may exist through different channels: transfer of tacit knowledge, more scientific publications, economy of scale in knowledge production because of better research equipment, and expanded research network. The results also verified the positive effect of funding, the positive effect of past networking (measured by number of co-authors), the inverted U-shaped effect of age, and the fewer number of publications by women compared to men.

Considering the significant and positive effect of collaboration with top-funded scientists on scientific productivity, the main policy advice is to continue to promote incentives for such collaborations and to maximize the benefits of such collaborations. This kind of collaboration may positively affect the academic productivity through different channels. As discussed above it may result in a substantial transfer of tacit knowledge and more scientific publications. Moreover, it provides benefits such as economy of scale in knowledge production because well-funded scientists have generally larger teams of researchers and better research equipment. The expansion of the research network is another reviewed benefit of such collaboration. Collaboration with top-funded scientists has also some interactive effects with other determinants of scientific production, which suggests an amplified positive impact if a researcher has a greater number of co-authors, if a researcher is female. In terms of interaction with private funding and NFP funding, there is not a systemic and structured pattern. However, some of interactive variables have positive effects.

The main purpose of chapter 5 was to make a distinction between different attributes of research chairs and their effect on scientific productivity. One of the important questions is to find out whether a research chair still has a better scientific productivity (compared to non-chair) after controlling the exclusive fund of research chair. To investigate that, we employed a matching technique to make pairs of scientists (chair and non-chair holders) with similar gender, funding and research field. After such matching, we have found that the effect of the Canada research chair program on scientific productivity remains significant and positive while the effect of industrial chairs and the chairs appointed by the Canadian federal granting council (NSERC, and CIHR) become non-significant. This finding highlights the effectiveness of our matching technique methodology because before matching holding any type of chair had positive and significant effect on scientific productivity.

This finding highlights the special attributes of the Canada research chair program, which are not replicated in other chairs. Those specific attributes may significantly push scientific productivity. Among others, Canada research chairs are generally associated with some degree of prestige or higher visibility to recruit talented students or to have research collaboration with top scientists in the field. The fact that other types of research chairs, once matched with equivalent scientists, do not have an impact on scientific output in terms of quantity, does not imply that these chair holders are lesser scientists, but that they are devoting part of their time to other endeavours of a more practical nature. Hence universities are maintaining a balance between the pursuit of pure scientific knowledge and its application to socio-economic benefits. We are not going to shock anyone by stating that by solely studying scientific articles, we are missing a great deal of the role of university professors. Although not trivial, future research should aim to cast a wider net on outputs, outcomes and impacts of university research.

This thesis brings some novel advancements in the analysis of scientists' productivity. In terms of data set we used, there is a mix of bibliometric and funding data in an econometric study. There are a limited number of scientific papers that investigate the effect of funding while controlling other bibliometric indices. Such comprehensive empirical analysis improve the reliability and significance of research results. Second, we have conducted a comprehensive investigation on determinants of research impact and citation count. To the best of our knowledge, there is no study that identifies the determinants of scientific productivity and research productivity of scientist in the province of Quebec using comprehensive econometric analysis. This thesis provides a relatively comprehensive study on this topic.

Third, we developed a theoretical model to predict the effect of collaboration with top-funded scientist and test the prediction with real data. Using contract theory and mechanism design modeling, this thesis tries to provide a theoretical prediction for empirical investigation. Last, we have used matching techniques to find out the influence of holding a research chair on scientific productivity: this technique improve the robustness and reliability of the result because it disentangles the prestige effect from the funding effect of the chair and hence, 'holding a research chair' becomes a better and more informative signal for the prestige of scientists.

There are some suggestions for future works on this subject. The first paper use the number of citation as measure of research impact. Other measures and proxies can be used in future studies to check whether determinants are still significant. The second paper defines the "star" scientists based on the amount of funding, i.e. it identifies the top-funded scientists. As the funding of scientist is being determined by previous productivity and intrinsic characteristics of scientists, it is an appropriate proxy for research impact. However, future research can consider different definitions of "star scientist" to provide a more comprehensive and robust interpretation. For instance, star scientists can be identified based on the number of patents, the impact factor of the journals in which the researchers publish, and networking measure such as centrality or cliquishness indicators. Another suggestion for future studies is about different research methods. One can have a deep investigation on the nature of research collaboration and provide a taxonomy to explain the benefit and knowledge spillover of different types of research collaboration with star scientists. This will require data gathering by surveys and questionnaires in order to have a detailed understating the nature of the collaboration with star scientists.

In terms of suggestion for future studies related to the third paper, we can note about qualitative analysis on the attributes of the research chairs. Gathering data from interviews with chair-holders or surveys filled up by experts would shed some light on how holding a research chair contributes to scientific productivity. Such qualitative work could help investigate the effect of research chair from a social or psychological point of view. Moreover, it would also be possible to find the institutional effect of the research chairs or the impact of peripheral institutional settings on the productivity of research chairs.

BIBLIOGRAPHY

Aaltojärvi, I., Arminen, I., Auranen, O., Pasanen, H.-M., 2008. Scientific productivity, web visibility and citation patterns in sixteen Nordic sociology departments. Acta Sociologica 51, 5-22.

Abramo, G., D'Angelo, C.A., Caprasecca, A., 2009a. Gender differences in research productivity: A bibliometric analysis of the Italian academic system. Scientometrics 79, 517-539.

Abramo, G., D'Angelo, C.A., Di Costa, F., 2009b. Research collaboration and productivity: is there correlation? Higher Education 57, 155-171.

Adam, D., 2002. Citation analysis: The counting house. Nature 415, 726-729.

Aksnes, D.W., 2006. Citation rates and perceptions of scientific contribution. Journal of the American Society for Information Science and Technology 57, 169-185.

Aksnes, D.W., Rorstad, K., Piro, F., Sivertsen, G., 2011. Are female researchers less cited? A large-scale study of Norwegian scientists. Journal of the American Society for Information Science and Technology 62, 628-636.

Ale Ebrahim, N., Salehi, H., Embi, M.A., Habibi, F., Gholizadeh, H., Motahar, S.M., 2014. Visibility and citation impact. International Education Studies 7, 120-125.

Allison, P.D., Long, J.S., Krauze, T.K., 1982. Cumulative advantage and inequality in science. American Sociological Review, 615-625.

Allison, P.D., Stewart, J.A., 1974. Productivity differences among scientists: Evidence for accumulative advantage. American Sociological Review, 596-606.

Amsterdamska, O., Leydesdorff, L., 1989. Citations: indicators of significance? Scientometrics 15, 449-471.

Appleyard, M.M., 1996. How does knowledge flow? Interfirm patterns in the semiconductor industry. Strategic management journal 17, 137-154.

Arundel, A., Geuna, A., 2004. Proximity and the use of public science by innovative European firms. Economics of Innovation and new Technology 13, 559-580.

Azoulay, P., Ding, W., Stuart, T., 2007. The determinants of faculty patenting behavior: Demographics or opportunities? Journal of Economic Behavior & Organization 63, 599-623.

Azoulay, P., Graff Zivin, J., Wang, J., 2010. Superstar extinction. Quarterly Journal of Economics 25, 549-589.

Bagchi-Sen, S., 2007. Strategic considerations for innovation and commercialization in the US biotechnology sector. European Planning Studies 15, 753-766.

Baird, L.L., 1986. What characterizes a productive research department? Research in Higher Education 25, 211-225.

Baird, L.L., 1991. Publication productivity in doctoral research departments: Interdisciplinary and intradisciplinary factors. Research in Higher Education 32, 303-318.

Balconi, M., Laboranti, A., 2006. University–industry interactions in applied research: The case of microelectronics. Research Policy 35, 1616-1630.

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., Terliesner, J., 2012. Beyond citations: Scholars' visibility on the social Web. arXiv preprint arXiv:1205.5611.

Beaudry, C., Schiffauerova, A., 2011. Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. European Management Journal 29, 362-376.

Becker, G.S., 1962. Investment in human capital: a theoretical analysis. The journal of political economy 70, 9-49.

Beise, M., Stahl, H., 1999. Public research and industrial innovations in Germany. Research Policy 28, 397-422.

Bercovitz, J., Feldman, M., 2011. The mechanisms of collaboration in inventive teams: composition, social networks, and geography. Research Policy 40, 81-93.

Berman, E.M., 1990. The economic impact of industry-funded university R&D. Research Policy 19, 349-355.

Bernier, C., Gill, W., Hunt, R., 1975. Measures of excellence of engineering and science departments: a chemical engineering example. Chemical Engineering Education 9, 194-202.

Bérubé, C., Mohnen, P., 2009. Are firms that receive R&D subsidies more innovative? Canadian Journal of Economics/Revue canadienne d'économique 42, 206-225.

Blackburn, R.T., Behymer, C.E., Hall, D.E., 1978. Research note: Correlates of faculty publications. Sociology of Education, 132-141.

Bonaccorsi, A., Daraio, C., 2003. Age effects in scientific productivity. Scientometrics 58, 49-90.

Bornmann, L., Daniel, H.-D., 2008. What do citation counts measure? A review of studies on citing behavior. Journal of Documentation 64, 45-80.

Bornmann, L., Mutz, R., Neuhaus, C., Daniel, H.-D., 2008. Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. Ethics in Science and Environmental Politics 8, 93-102.

Broad, W.J., 1981. The publishing game: getting more for less. Science 211, 1137-1139.

Brown, L.D., Gardner, J.C., 1985. Applying citation analysis to evaluate the research contributions of accounting faculty and doctoral programs. Accounting Review, 262-277.

Buchmueller, T.C., Dominitz, J., Lee Hansen, W., 1999. Graduate training and the early career productivity of Ph.D. economists. Economics of Education Review 18, 65-77.

Butler, L., 2002. A list of published papers is no measure of value. Nature 419, 877-877.

Butler, L., 2003a. Explaining Australia's increased share of ISI publications--the effects of a funding formula based on publication counts. Research Policy 32, 143-155.

Butler, L., 2003b. Modifying publication practices in response to funding formulas. Research Evaluation 12, 39-46.

Butler, L., 2007. Assessing university research: a plea for a balanced approach. Science and Public Policy 34, 565-574.

Calderini, M., Franzoni, C., 2004. Is academic patenting detrimental to high quality research. An empirical analysis of the relationship between scientific careers and patent applications. Bocconi University: Cespri Working Paper.

Callon, M., 1994. Is Science a Public Good? Fifth Mullins Lecture, Virginia Polytechnic Institute, 23 March 1993. Science, Technology & Human Values 19, 395-424.

Cantu, F.J., Bustani, A., Molina, A., Moreira, H., 2009. A knowledge-based development model: the research chair strategy. Journal of Knowledge Management 13, 154-170.

Carayol, N., Matt, M., 2004. Does research organization influence academic production?:: Laboratory level evidence from a large European university. Research Policy 33, 1081-1102.

Carayol, N., Matt, M., 2006. Individual and collective determinants of academic scientists' productivity. Information Economics and Policy 18, 55-72.

Carew, R., 2001. Institutional arrangements and public agricultural research in Canada. Review of agricultural economics 23, 82-101.

Chwe, M.S.Y., 2000. Communication and coordination in social networks. Review of economic studies 67, 1-16.

Cole, J., Cole, S., 1971. Measuring the Quality of Sociological Research: Problems in the Use of the" Science Citation Index". The American Sociologist 6, 23-29.

Courty, P., Sim, J., 2012. What is the cost of retaining and attracting exceptional talents? Evidence from the Canada Research Chair program. Queen's Economics Department Working Paper.

Crane, D., 1965. Scientists at major and minor universities: A study of productivity and recognition. American Sociological Review, 699-714.

Crespi, G.A., Geuna, A., 2008. An empirical study of scientific production: A cross country analysis, 1981–2002. Research Policy 37, 565-579.

Darby, M.R., Zucker, L.G., 1996. Star scientists, institutions, and the entry of Japanese biotechnology enterprises. National bureau of economic research.

Diamond, A.M., 1986. The life-cycle research productivity of mathematicians and scientists. Journal of Gerontology 41, 520.

Durnev, A., Kim, E., 2005. To steal or not to steal: Firm attributes, legal environment, and valuation. The Journal of Finance 60, 1461-1493.

Favaloro, E.J., 2008. Measuring the quality of journals and journal articles: the impact factor tells but a portion of the story. New York: Stratton Intercontinental Medical Book Corporation, c1974-, pp. 7-25.

Feist, G.J., 1997. Quantity, quality, and depth of research as influences on scientific eminence: Is quantity most important? Creativity Research Journal 10, 325-335.

Feldman, M.P., Graddy-Reed, A., 2013. Accelerating commercialization: a new model of strategic foundation funding. The Journal of Technology Transfer, 1-21.

Fisher, R.S., Powers, L.E., 2004. Peer-Reviewed Publication: A View from Inside. Epilepsia 45, 889-894.

Fowler, J.H., Aksnes, D.W., 2007. Does self-citation pay? Scientometrics 72, 427-437.

Fox, M.F., 2005. Gender, family characteristics, and publication productivity among scientists. Social Studies of Science 35, 131-150.

Franceschet, M., 2010. The difference between popularity and prestige in the sciences and in the social sciences: A bibliometric analysis. Journal of Informetrics 4, 55-63.

Frey, B.S., Rost, K., 2010. Do rankings reflect research quality? Journal of Applied Economics 13, 1-38.

Geuna, A., 2001. The changing rationale for European university research funding: are there negative unintended consequences? Journal of Economic Issues 35, 607-632.

Geuna, A., Nesta, L., 2003. University patenting and its effects on academic research. SEWPS Paper.

Godin, B., Gingras, Y., 2000. Impact of collaborative research on academic science. Science and Public Policy 27, 65-73.

Golden, J., Carstensen, F.V., 1992. Academic research productivity, department size and organization: Further results, comment. Economics of Education Review 11, 153-160.

Goldfarb, B., 2008. The effect of government contracting on academic research: Does the source of funding affect scientific output? Research Policy 37, 41-58.

Gonzalez-Brambila, C., Veloso, F.M., 2007. The determinants of research output and impact: A study of Mexican researchers. Research Policy 36, 1035-1051.

Groot, T., Garcia-Valderrama, T., 2006. Research quality and efficiency: an analysis of assessments and management issues in Dutch economics and business research programs. Research Policy 35, 1362-1376.

Gulbrandsen, M., Smeby, J.C., 2005. Industry funding and university professors' research performance. Research Policy 34, 932-950.

Handberg, R., 1986. Practising Western Science Inside the West: Psychological and Institutional Parallels between Western and Nonwestern Academic Cultures. Social Studies of Science 16, 529-534.

Harman, G., 2000. Allocating research infrastructure grants in post-binary higher education systems: British and Australian approaches. Journal of Higher Education Policy and Management 22, 111-126.

Hayati, Z., Ebrahimy, S., 2009. Correlation between quality and quantity in scientific production: A case study of Iranian organizations from 1997 to 2006. Scientometrics 80, 625-636.

Hayes, S.C., 1983. When more is less: Quantity versus quality of publications in the evaluation of academic vitae. American Psychologist 38, 1398.

Heinze, T., Shapira, P., Rogers, J.D., Senker, J.M., 2009. Organizational and institutional influences on creativity in scientific research. Research Policy 38, 610-623.

Henrekson, M., Waldenström, D., 2007. How should research performance be measured. IFN Working Paper.

Hesli, V.L., Lee, J.M., 2011. Faculty Research Productivity: Why Do Some of Our Colleagues Publish More than Others? PS Political Science and Politics 44, 393.

Hicks, D., 1995. Published papers, tacit competencies and corporate management of the public/private character of knowledge. Industrial and corporate change 4, 401-424.

Higgins, M.J., Stephan, P.E., Thursby, J.G., 2011. Conveying quality and value in emerging industries: Star scientists and the role of signals in biotechnology. Research Policy 40, 605-617.

Ho, W., Dey, P.K., Higson, H.E., 2006. Multiple criteria decision-making techniques in higher education. International journal of educational management 20, 319-337.

Hottenrott, H., Thorwarth, S., 2011. Industry funding of university research and scientific productivity. Kyklos 64, 534-555.

Jaffe, A.B., 1989. Real effects of academic research. The American Economic Review, 957-970.

Jensen, M.C., Meckling, W.H., 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of financial economics 3, 305-360.

Johnes, G., 1988. Determinants of research output in economics departments in British universities. Research Policy 17, 171-178.

Jordan, J.M., Meador, M., Walters, S.J.K., 1988. Effects of department size and organization on the research productivity of academic economists. Economics of Education Review 7, 251-255.

Jordan, J.M., Meador, M., Walters, S.J.K., 1989. Academic research productivity, department size and organization: Further results. Economics of Education Review 8, 345-352.

Jungnickel, P.W., Creswell, J.W., 1994. Workplace correlates and scholarly performance of clinical pharmacy faculty. Research in Higher Education 35, 167-194.

Katz, J.S., Martin, B.R., 1997. What is research collaboration? Research Policy 26, 1-18.

Kleinman, D.L., Vallas, S.P., 2001. Science, capitalism, and the rise of the "knowledge worker": The changing structure of knowledge production in the United States. Theory and Society 30, 451-492.

Kostoff, R.N., 1998. The use and misuse of citation analysis in research evaluation. Scientometrics 43, 27-43.

Kostoff, R.N., 2002. Citation analysis of research performer quality. Scientometrics 53, 49-71.

Kyvik, S., 1990. Age and scientific productivity. Differences between fields of learning. Higher Education 19, 37-55.

Kyvik, S., 1995. Are big university departments better than small ones? Higher Education 30, 295-304.

Kyvik, S., Olsen, T.B., 2008. Does the aging of tenured academic staff affect the research performance of universities? Scientometrics 76, 439-455.

Kyvik, S., Teigen, M., 1996. Child care, research collaboration, and gender differences in scientific productivity. Science, Technology & Human Values 21, 54.

Laband, D.N., Piette, M.J., 1994. The relative impacts of economics journals: 1970-1990. Journal of Economic Literature 32, 640-666.

Leahey, E., 2006a. Gender differences in productivity. Gender & Society 20, 754-780.

Leahey, E., 2006b. Gender Differences in Productivity Research Specialization as a Missing Link. Gender & Society 20, 754-780.

Lertputtarak, S., 2008. An investigation of factors related to research productivity in a public university in Thailand: a case study. Victoria University.

Levin, S.G., Stephan, P.E., 1991. Research productivity over the life cycle: evidence for academic scientists. The American Economic Review, 114-132.

Liefner, I., 2003. Funding, resource allocation, and performance in higher education systems. Higher Education 46, 469-489.

Long, J.S., 1990. The origins of sex differences in science. Social Forces 68, 1297-1316.

Long, J.S., 1992. Measures of sex differences in scientific productivity. Social Forces 71, 159-178.

Long, J.S., Allison, P.D., McGinnis, R., 1979. Entrance into the academic career. American Sociological Review, 816-830.

Luukkonen, T., Tijssen, R.J.W., Persson, O., Sivertsen, G., 1993. The measurement of international scientific collaboration. Scientometrics 28, 15-36.

Manjarrés-Henríquez, L., Gutiérrez-Gracia, A., Carrión-García, A., Vega-Jurado, J., 2009. The effects of university–industry relationships and academic research on scientific performance: Synergy or substitution? Research in Higher Education 50, 795-811.

Manjarrés-Henríquez, L., Gutiérrez-Gracia, A., Vega-Jurado, J., 2008. Coexistence of university-industry relations and academic research: Barrier to or incentive for scientific productivity. Scientometrics 76, 561-576.

Meho, L.I., 2007. The rise and rise of citation analysis. Physics World.

Melin, G., 1996. The networking university. Scientometrics 35, 15-31.

Melin, G., Persson, O., 1996. Studying research collaboration using co-authorships. Scientometrics 36, 363-377.

Merton, R.K., 1968. The Matthew effect in science. Science 159, 56-63.

Meyer, M., 2006. Are patenting scientists the better scholars?: An exploratory comparison of inventor-authors with their non-inventing peers in nano-science and technology. Research Policy 35, 1646-1662.

Moed, H.F., 2005. Citation analysis in research evaluation. Springer.

Moed, H.F., 2006. Citation analysis in research evaluation. Springer.

Moed, H.F., 2009. New developments in the use of citation analysis in research evaluation. Archivum immunologiae et therapiae experimentalis 57, 13-18.

Moed, H.F., Burger, W., Frankfort, J., Van Raan, A.F.J., 1985. The use of bibliometric data for the measurement of university research performance. Research Policy 14, 131-149.

Moed, H.F., Van Leeuwen, T.N., 1995. Improving the accuracy of Institute for Scientific Information's journal impact factors. Journal of the American Society for Information Science 46, 461-467.

Nakhaie, M.R., 2002. Gender Differences in Publication among University Professors in Canada*. Canadian Review of Sociology/Revue canadienne de sociologie 39, 151-179.

Narin, F., 1976. Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity. Computer Horizons Washington, D. C.

Narin, F., Hamilton, K.S., Olivastro, D., 1997. The increasing linkage between US technology and public science. Research Policy 26, 317-330.

Neumann, Y., Finaly-Neumann, E., 1990. The support-stress paradigm and faculty research publication. The Journal of Higher Education, 565-580.

Niosi, J., 2002. Regional systems of innovation: Market pull and government push. Holbrook, J.-A and D. Wolfe,(eds.) Knowledge, Clusters and Regional Innovation. Montréal & Kingston, McGill-Queen's University Press, 39-55.

Niosi, J., Bas, T.G., 2004. Canadian biotechnology policy: designing incentives for a new technology. Environment and Planning C 22, 233-248.

Norris, M., Oppenheim, C., 2003. Citation counts and the Research Assessment Exercise V: Archaeology and the 2001 RAE. Journal of Documentation 59, 709-730.

Oettl, A., 2009. Productivity and helpfulness: Implications of a new taxonomy for star scientists. mimeo Georgia Tech.

Olson, C.M., Rennie, D., Cook, D., Dickersin, K., Flanagin, A., Hogan, J.W., Zhu, Q., Reiling, J., Pace, B., 2002. Publication bias in editorial decision making. JAMA: the journal of the American Medical Association 287, 2825-2828.

Partha, D., David, P.A., 1994. Toward a new economics of science. Research Policy 23, 487-521.

Pavitt, K., 1991. What makes basic research economically useful? Research Policy 20, 109-119.

Pavitt, K., 2000. Why European Union funding of academic research should be increased: a radical proposal. Science and Public Policy 27, 455-460.

Pavitt, K., 2001. Public policies to support basic research: What can the rest of the world learn from US theory and practice?(And what they should not learn). Industrial and corporate change 10, 761-779.

Phelan, T., 1999. A compendium of issues for citation analysis. Scientometrics 45, 117-136.

Phillips, P.W.B., 2010. Genomics and public policy: Wealth for Canadians. Integrated Assessment 10.

Radicchi, F., Fortunato, S., Markines, B., Vespignani, A., 2009. Diffusion of scientific credits and the ranking of scientists. Physical Review E 80, 056103.

Ravenscraft, D., Scherer, F.M., 1982. The lag structure of returns to research and development. Applied Economics 14, 603-620.

Rothaermel, F.T., Hess, A.M., 2007. Building dynamic capabilities: Innovation driven by individual-, firm-, and network-level effects. Organization Science 18, 898-921.

Salazar, M., Holbrook, A., 2007. Canadian science, technology and innovation policy: the product of regional networking? Regional studies 41, 1129-1141.

Salter, A.J., Martin, B.R., 2001. The economic benefits of publicly funded basic research: a critical review. Research Policy 30, 509-532.

Sammarco, P.W., 2008. Journal visibility, self-citation, and reference limits: influences on Impact Factor and author performance review. Ethics in Science and Environmental Politics 8, 121-125.

Schiffauerova, A., Beaudry, C., 2011. Star scientists and their positions in the Canadian biotechnology network. Economics of Innovation and new Technology 20, 343-366.

Schimank, U., 2005. 'New Public Management'and the academic profession: Reflections on the German situation. Minerva 43, 361-376.

Seglen, P.O., 1994. Causal relationship between article citedness and journal impact. Journal of the American Society for Information Science 45, 1-11.

Seglen, P.O., 1997a. Citations and journal impact factors: questionable indicators of research quality. Allergy 52, 1050-1056.

Seglen, P.O., 1997b. Why the impact factor of journals should not be used for evaluating research. Bmj 314, 497.

Smeby, J.C., Try, S., 2005. Departmental contexts and faculty research activity in Norway. Research in Higher Education 46, 593-619.

Stegmann, J., Grohmann, G., 2001. Citation rates, knowledge export and international visibility of dermatology journals listed and not listed in theJournal Citation Reports. Scientometrics 50, 483-502.

Sugimoto, C.R., Lariviere, V., Ni, C., Gingras, Y., Cronin, B., 2013. Global gender disparities in science. Nature 504, 211-213.

Tijssen, R.J.W., 2004. Is the commercialisation of scientific research affecting the production of public knowledge?: Global trends in the output of corporate research articles. Research Policy 33, 709-733.

Turner, L., Mairesse, J., 2005. Individual Productivity Differences in Public Research: How important are non-individual determinants? An Econometric Study of French Physicists' publications and citations (1986-1997). Centre National de la Recherche Scientifique.

Van Raan, A.F.J., 2005. Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. Scientometrics 62, 133-143.

West, M.A., Smith, H., Feng, W.L., Lawthom, R., 1998. Research excellence and departmental climate in British universities. Journal of Occupational and Organizational Psychology 71, 261-281.

Winter, M., Smith, C., Morris, P., Cicmil, S., 2006. Directions for future research in project management: The main findings of a UK government-funded research network. International Journal of Project Management 24, 638-649.

Xie, Y., Shauman, K.A., 1998. Sex differences in research productivity: New evidence about an old puzzle. American Sociological Review, 847-870.

Ynalvez, M.A., Shrum, W.M., 2011. Professional networks, scientific collaboration, and publication productivity in resource-constrained research institutions in a developing country. Research Policy 40, 204-216.

Zhou, J., 2003. When the presence of creative coworkers is related to creativity: role of supervisor close monitoring, developmental feedback, and creative personality. Journal of Applied Psychology 88, 413.

Zhou, Y.B., Lü, L., Li, M., 2012. Quantifying the influence of scientists and their publications: distinguishing between prestige and popularity. New Journal of Physics 14, 033033.

Zucker, L.G., Darby, M.R., 1996. Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. Proceedings of the National Academy of Sciences 93, 12709-12716.

Zucker, L.G., Darby, M.R., 1997. Individual Action and the Demand for Institutions Star Scientists and Institutional Transformation. American Behavioral Scientist 40, 502-513.

Zucker, L.G., Darby, M.R., 2001. Capturing technological opportunity via Japan's star scientists: Evidence from Japanese firms' biotech patents and products. The Journal of Technology Transfer 26, 37-58.

Zucker, L.G., Darby, M.R., Armstrong, J.S., 2002. Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology. Management Science 48, 138-153.

APPENDICES

APPENDIX A – VARIABLE DESCRIPTION OF CHAPTER 3

Variable name	Variable description
ln(*nbCitation*)$_{it}$	Natural logarithm of number of citations of papers published by scientist i in year t (10 years following publication year) divided by the average citation rate of the papers published in the same year in the same discipline
ln(*PublicfundingO*)$_{it}$	Natural logarithm of the three-year average up to year t of public sector funding for the purpose of operational costs and direct expenditures of research of researcher i
ln(*PublicfundingI*)$_{it}$	Natural logarithm of the three-year average up to year t of public sector funding for the purpose of buying instruments for researcher i
ln(*PrivatefundingO*)$_{it}$	Natural logarithm of the three-year average up to year t of private sector funding for the purpose of operational costs and direct expenditures of research of researcher i
ln(*NFPfundingO*)$_{it}$	Natural logarithm of three-year average up to year t of funding from not-for-profit institutions (NFP) for the purpose of operational costs and direct expenditures of research of researcher i
ln(*nbArticle*)$_{it}$	Natural logarithm of number of articles published in year t by researcher i
ln(*nbAuthor*)$_{it}$	Natural logarithm of the three-year average up to year t of number of authors in the papers of researcher i
ln(*Impactfactor*)$_{it}$	Natural logarithm of the five-year average up to year t of journal impact factor in which the scientist publishes
ln(*nbArticle*)$_{it}$* ln(*Impactfactor*)$_{it}$	Interaction between ln(*nbArticle*)$_{it}$ and ln(*Impactfactor*)$_{it}$: ln(*nbArticle*)$_{it}$ X ln(*Impactfactor*)$_{it}$
dFemalei	Dummy variable taking the value 1 if the scientist is a woman and 0 otherwise
Ageit	Age of a researcher i at year t
d2000, d2001, ….	Dummy variables indicating the year
ln(nbArticleAvg3)$_{it}$	Natural logarithm of the three-year average up to year t of articles published by researcher i

APPENDIX B – SUMMARY STATISTICS OF CHAPTER 3

variable	*Average*	*Standard Deviation*	*Minimum*	*Maximum*
nbCitation	1.183177	1.74	0	74.575
Age	50.62187	9.26	16	92
dfemale	0.228939	0.42	0	1
nbArticle	3.469125	3.66	0	84
Impactfactor	1.130298	0.63	0.016	12.476
nbAuthor	7.269147	52.02	1	3174.5
PublicfundingO	112969.6	203169.30	0	1.01E+07
PrivatefundingO	21543.06	96437.42	0	4077193
NFPfundingO	21294.22	128658.50	0	8720387
PublicfundingI	30482.68	222562.50	0	1.28E+07

The variables are not summarized in logarithmic scale and they are raw amount. In some disciplines of Physics, there are many scientists involved in one project and therefore, the maximum for the number of authors is high. (No. observation: = 31,563)

APPENDIX C – CORRELATION TABLE OF CHAPTER 3

		1	2	3	4	5	6	7	8	9	10	11	12
ln(*nbCitation*)$_{it}$	**1**	1											
ln(*PublicfundingI*)$_{it-1}$	**2**	0.0733	1										
Age$_{it}$	**3**	-0.0377	-0.1247	1									
ln(*nbArticleAvg3*)$_{it-1}$	**4**	0.1945	0.2000	0.1130	1								
dFemale$_{i}$	**5**	-0.0318	-0.1020	-0.1197	-0.1203	1							
ln(*nbArticle*)$_{it}$	**6**	0.2359	0.1912	0.0136	0.6728	-0.0995	1						
ln(*Impactfactor*)$_{it}$	**7**	0.4686	0.1029	-0.0440	0.2209	-0.0406	0.2192	1					
ln(*nbArticle*)$_{it}$*ln(*Impactfactor*)$_{it}$	**8**	0.4695	0.1114	-0.0409	0.2772	-0.0449	0.3005	0.9087	1				
ln(*nbAuthor*)$_{it}$	**9**	0.2724	0.0039	0.0668	0.3098	0.0046	0.2898	0.2382	0.2794	1			
ln(*PublicfundingO*)$_{it}$	**10**	0.0707	0.2905	-0.1044	0.2571	-0.0218	0.2347	0.1094	0.1163	-0.0002	1		
ln(*PrivatefundingO*)$_{it}$	**11**	0.0723	0.1415	-0.007	0.2141	-0.1219	0.2163	0.0487	0.0550	0.1385	0.1294	1	
ln(*NFPfundingO*)$_{it}$	**12**	0.1017	0.1145	-0.0603	0.2470	-0.0178	0.2339	0.1216	0.1422	0.1952	0.1963	0.2227	1

Significant at 1% level (No. observation: 31,563)

APPENDIX D – FIRST STAGE REGRESSION RESULTS OF CHAPTER 3

Dependent variable: *PublicfundingO*	**Domain**												
	A	**B**	**A+B**	**C**	**D**	**C+D**	**E**	**F**	**G**	**H**	**I**	**H+I**	**All**
***dFemale**$_i$*	0.2876 ***	0.2993 *	0.4328 ***	0.2781	0.2907 **	0.2834 ***	-0.1251	0.1731	0.9591 **	-0.0572	-0.0723	-0.0683	0.1568 ***
	(0.1103)	(0.1687)	(0.0938)	(0.2072)	(0.1213)	(0.1064)	(0.4732)	(0.2831)	(0.3760)	(0.1405)	(0.0996)	(0.0806)	(0.0492)
ln(*nbArticle*)$_{it}$	0.4649 ***	0.3078 **	0.4427 ***	0.7069 ***	0.4027 ***	0.4487 ***	0.4014	0.3092	0.2613	0.1989 ***	0.3317 ***	0.2856 ***	0.3922 ***
	(0.0815)	(0.1534)	(0.0725)	(0.2701)	(0.1174)	(0.1077)	(0.5902)	(0.5991)	(0.3978)	(0.0667)	(0.0562)	(0.0429)	(0.0343)
ln(*Impactfactor*)$_{it}$	0.6793 ***	0.1979	0.5225 ***	0.1001	0.2232	0.2232	0.2278	0.1913	0.4793	0.2884 **	0.4161 ***	0.3557 ***	0.3085 ***
	(0.1670)	(0.3367)	(0.1496)	(0.4328)	(0.2017)	(0.1796)	(0.6987)	(0.6751)	(0.6336)	(0.1364)	(0.1216)	(0.0905)	(0.0658)
ln(*nbArticle*)$_{it}$*ln(*Impactfactor*)$_{it}$	-0.2200	-0.1469	-0.1574	0.2390	-0.0709	-0.0426	-0.2590	-0.0838	-0.3334	-0.1483	-0.2412 **	-0.1954 **	-0.1326 **
	(0.1394)	(0.3063)	(0.1265)	(0.4913)	(0.2157)	(0.1947)	(0.8194)	(0.8580)	(0.7133)	(0.1231)	(0.1044)	(0.0792)	(0.0594)
ln(*nbAuthor*)$_{it}$	-0.4378 ***	-0.1446	-0.4527 ***	-0.2270	0.1611 *	0.1010	0.6436 *	0.3161	-0.6015 **	-0.0372	-0.2073 ***	-0.1671 ***	-0.3551 ***
	(0.0710)	(0.1445)	(0.0639)	(0.2269)	(0.0912)	(0.0847)	(0.3632)	(0.2572)	(0.2847)	(0.0716)	(0.0456)	(0.0375)	(0.0286)
ln(*PrivatefundingO*)$_{it}$	-0.0382 ***	0.0556 ***	-0.0318 ***	0.1253 ***	0.0663 ***	0.0893 ***	0.0786	0.1491 *	0.0189	0.0962 ***	0.0676 ***	0.0776 ***	0.0267 ***
	(0.0082)	(0.0184)	(0.0075)	(0.0261)	(0.0191)	(0.0153)	(0.0887)	(0.0779)	(0.0475)	(0.0071)	(0.0068)	(0.0049)	(0.0038)
ln(*NFPfundingO*)$_{it}$	0.0895 ***	0.0451 ***	0.0800 ***	0.0897 ***	0.0490 ***	0.0568 ***	0.1222 ***	0.1051 *	0.1543 ***	0.0489 ***	0.0601 ***	0.0560 ***	0.0574 ***
	(0.0077)	(0.0157)	(0.0070)	(0.0247)	(0.0131)	(0.0116)	(0.0562)	(0.0589)	(0.04210	(0.0080)	(0.0066)	(0.0051)	(0.0036)
ln(*PublicfundingI*)$_{it-1}$	0.1217 ***	0.1007 ***	0.1191 ***	0.0719 ***	0.0769 ***	0.0728 ***	0.0089	0.0837 *	0.1662 ***	0.0666 ***	0.0754 ***	0.0717 ***	0.0927 ***
	(0.0088)	(0.0209)	(0.0081)	(0.0244)	(0.0129)	(0.0113)	(0.0631)	(0.0451)	(0.0568)	(0.0065)	(0.0058)	(0.0043)	(0.0036)
Age$_{it}$	0.1304 ***	0.4570 ***	0.1872 ***	0.1685 *	0.0468	0.0370	0.5601 ***	-0.1593	0.3943 **	0.1431 ***	0.2126 ***	0.1886 ***	0.1550 ***
	(0.0399)	(0.0823)	(0.0359)	(0.0951)	(0.0474)	(0.0425)	(0.1889)	(0.1288)	(0.1627)	(0.0331)	(0.0296)	(0.0223)	(0.0171)
Age$^2_{it}$	-0.0017 ***	-0.0048 ***	-0.0022 ***	-0.0023 ***	-0.0007	-0.0007	-0.0061 ***	0.0014	-0.0041 **	-0.0016 ***	-0.0024 ***	-0.0021 ***	-0.0018 ***
	(0.0004)	(0.0008)	(0.0003)	(0.0010)	(0.0005)	(0.0004)	(0.0018)	(0.0013)	(0.0016)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
ln(*nbArticleAavg3*)$_{it-1}$	1.5659 ***	0.8759 ***	1.4303 ***	1.1544 ***	1.1444 ***	1.1767 ***	0.8231 *	1.2543 ***	-0.1928	0.6174 ***	0.7661 ***	0.7135 ***	0.9977 ***
	(0.0816)	(0.1489)	(0.0725)	(0.2222)	(0.1085)	(0.0969)	(0.4539)	(0.4056)	(0.3582)	(0.0663)	(0.0568)	(0.0432)	(0.0339)
Constant	4.4273 ***	-2.4294 0.2430	3.2488 ***	4.9033 **	6.9708 ***	7.3620 ***	-6.1381	11.3988 ***	-0.3949	5.7952 ***	4.0233 ***	4.6705 ***	4.9276 ***
	(1.0653)	(2.0788)	(0.9517)	(2.3117)	(1.1890)	(1.0572)	(4.9201)	(3.2269)	(4.0890)	(0.8470)	(0.7648)	(0.5738)	(0.4411)
Number of observations	9026	1761	10787	1071	2944	4015	253	369	547	4029	7261	11290	31563
$\chi2$	1770 ***	256 ***	1906 ***	277 ***	510 ***	704 ***	75 ***	63 ***	124 ***	983 ***	1490 ***	2398 ***	4942 ***

*, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively – Year dummies and university dummies are significant· The definition of regression tags indicating the sample: A= Medical, B= Health Science, c= Business and Management, D= Social Science, E= Education, F= Humanities, G= Non-health professions, H= Engineering, I= Science.

APPENDIX E – OLS REGRESSION RESULTS OF CHAPTER 3

Dependent variable*: ln*(*nbCitation*)	**Domain**												
	A	**B**	**A+B**	**C**	**D**	**C+D**	**E**	**F**	**G**	**H**	**I**	**H+I**	**All**
dFemale$_{i}$	0.0072	0.0001	0.0074	0.0229	0.0103	0.0112	-0.0644	-0.1348 *	-0.0498	0.0067	-0.0245	-0.0161	-0.0049
	(0.0123)	(0.0201)	(0.0106)	(0.0389)	(0.0212)	(0.0186)	(0.0824)	(0.0732)	(0.0389)	(0.0297)	(0.0161)	(0.0141)	(0.0070)
ln(*nbArticle*)$_{it}$	0.0359 ***	0.0755 ***	0.0395 ***	0.0981 **	0.0817 ***	0.0812 ***	-0.0399	0.0416	0.0629	0.1203 ***	0.1037 ***	0.1083 ***	0.0667 ***
	(0.0088)	(0.0186)	(0.0080)	(0.0393)	(0.0202)	(0.0178)	(0.0853)	(0.1732)	(0.0416)	(0.0150)	(0.0134)	(0.0104)	(0.0057)
ln(*Impactfactor*)$_{it}$	0.2978 ***	0.2194 ***	0.2801 ***	0.3672 ***	0.2582 ***	0.2928 ***	0.2082	0.4411 **	0.0784	0.2249 ***	0.2626 ***	0.2497 ***	0.2461 ***
	(0.0278)	(0.0444)	(0.0248)	(0.0687)	(0.0392)	(0.0328)	(0.1367)	(0.1770)	(0.0775)	(0.0297)	(0.0383)	(0.0284)	(0.0154)
ln(*nbArticle*)$_{it}$***ln(*Impactfactor*)**$_{it}$	0.1172 ***	0.1185 ***	0.1201 ***	-0.0227	0.0630	0.0362	0.2107	-0.0940	0.1817 **	0.0641 **	0.0899 ***	0.0800 ***	0.1100 ***
	(0.0220)	(0.0388)	(0.0199)	(0.0781)	(0.0412)	(0.0350)	(0.1557)	(0.2287)	(0.0811)	(0.0256)	(0.0347)	(0.0259)	(0.0137)
ln(*nbAuthor*)$_{it}$	0.1944 ***	0.0499 *	0.1782 ***	0.0420	0.0802 ***	0.0662 ***	0.0186	0.0596	0.1710 ***	0.1440 ***	0.0728 ***	0.0892 ***	0.1090 ***
	(0.0141)	(0.0279)	(0.0131)	(0.0375)	(0.0187)	(0.0167)	(0.0712)	(0.0562)	(0.0465)	(0.0214)	(0.0177)	(0.0156)	(0.0086)
ln(*PublicfundingO*)$_{it}$	-0.0008	-0.0001	-0.0010	0.0085 *	0.0092 ***	0.0080 ***	0.0021	0.0129	-0.0014	0.0030	0.0040	0.0040	0.0001
	(0.0013)	(0.0029)	(0.0012)	(0.0046)	(0.0031)	(0.0025)	(0.0114)	(0.0096)	(0.0059)	(0.0036)	(0.0030)	(0.0024)	(0.0009)
ln(*PrivatefundingO*)$_{it}$	0.0043 ***	-0.0021	0.0036 ***	-0.0043	0.0040	0.0022	0.0126	-0.0262 ***	-0.0048	-0.0006	-0.0013	-0.0015	0.0013 **
	(0.0010)	(0.0025)	(0.0010)	(0.0053)	(0.0033)	(0.0029)	(0.0138)	(0.0092)	(0.0053)	(0.0018)	(0.0015)	(0.0011)	(0.0007)
ln(*NFPfundingO*)$_{it}$	-0.0030 ***	0.0031	-0.0021 ***	0.0072	0.0041	0.0050 **	0.0075	-0.0148	0.0066	-0.0006	0.0013	0.0010	-0.0009
	(0.0010)	(0.0024)	(0.0009)	(0.0044)	(0.0025)	(0.0022)	(0.0100)	(0.0118)	(0.0054)	(0.0019)	(0.0015)	(0.0012)	(0.0006)
Constant	0.2768 ***	0.3968 ***	0.2945 ***	0.2945 ***	0.2575 ***	0.2822 ***	0.8582 ***	0.4314 **	0.3331 **	0.1809 ***	0.3295 ***	0.2872 ***	0.3584 ***
	(0.0290)	(0.0696)	(0.0269)	(0.0820)	(0.0463)	(0.0398)	(0.2167)	(0.1896)	(0.1301)	(0.0494)	(0.0429)	(0.0352)	(0.0175)
Number of observations	10124	1954	12078	1243	3265	4508	281	420	613	4466	8089	12555	35201
Log likelihood	-4776.06	-825.66	-5675.60	-747.64	-1753.85	-2526.03	-193.77	-390.39	-345.74	-2162.48	-4108.04	-6312.27	-17887.10
F-test	.	25.34 ***	98.82 ***	21.31 ***	36.59 ***	53.31 ***	12.50 ***	57.48 ***	25.17 ***	45.19 ***	52.96 ***	84.53 ***	268.80 ***
R^2	0.32	0.25	0.31	0.28	0.26	0.26	0.33	0.28	0.28	0.22	0.24	0.24	0.27

*, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively – Year dummies and university dummies are significant. The definition of regression tags indicating the sample: A= Medical, B= Health Science, c= Business and Management, D= Social Science, E= Education, F= Humanities, G= Non-health professions, H= Engineering, I= Science.

APPENDIX F – PANEL REGRESSION OF CHAPTER 3 (no endogeneity assumption)

Dependent variable*: ln(nbCitation)*	**Domain**												
	A	**B**	**A+B**	**C**	**D**	**C+D**	**E**	F	G	H	I	H+I	All
dFemale$_i$	0.0061	0.0010	0.0058	0.0070	0.0100	0.0102	-0.0733	-0.1126	-0.0612	0.0009	-0.0268	-0.0183	-0.0063
	(0.0137)	(0.0228)	(0.0121)	(0.0375)	(0.0201)	(0.0182)	(0.0918)	(0.0708)	(0.0473)	(0.0311)	(0.0179)	(0.0157)	(0.0083)
ln(*nbArticle*)$_{it}$	0.0234 ***	0.0752) ***	0.0288 ***	0.0231	0.0689 ***	0.0560 ***	-0.0466	0.0425	0.0819 *	0.0753	0.0741 ***	0.0732 ***	0.0438 ***
	(0.0087)	(0.0189)	(0.0080)	(0.0396)	(0.0190)	(0.0171)	(0.1068)	(0.1327)	(0.0479)	(0.0130)	(0.0098)	(0.0077)	(0.0049)
ln(*Impactfactor*)$_{it}$	0.3027 ***	0.2358 ***	0.2889 ***	0.3958 ***	0.2709 ***	0.3049 ***	0.2484 *	0.4423 ***	0.1064	0.2359 ***	0.2380 ***	0.2346 ***	0.2472 ***
	(0.0193)	(0.0446)	(0.0177)	(0.0640)	(0.0342)	(0.0299)	(0.1346)	(0.1628)	(0.0798)	(0.0280)	(0.0223)	(0.0173)	(0.0099)
ln(*nbArticle*)$_{it}$*ln(*Impactfactor*)$_{it}$	0.0956 ***	0.0928 **	0.0958 ***	-0.0771	0.0372	0.0081	0.1043	-0.0995	0.1388	0.0523 **	0.0782 ***	0.0704 ***	0.0859 ***
	(0.0163)	(0.0405)	(0.0151)	(0.0728)	(0.0369)	(0.0327)	(0.1583)	(0.2059)	(0.0918)	(0.0256)	(0.0193)	(0.0152)	(0.0090)
ln(*nbAuthor*)$_{it}$	0.2045 ***	0.0601 ***	0.1884 ***	0.0528	0.0853 ***	0.0751 ***	0.0308	0.0674	0.1675 ***	0.1425 ***	0.1124 ***	0.1239 ***	0.1376 ***
	(0.0082)	(0.0195)	(0.0076)	(0.0339)	(0.0151)	(0.0139)	(0.0712)	(0.0573)	(0.0352)	(0.0151)	(0.0083)	(0.0072)	(0.0044)
ln(*PublicfundingO*)$_{it}$	-0.0006	-0.0001	-0.0008	0.0089 *	0.0073 **	0.0069 ***	0.0060	0.0120	-0.0039	0.0027	0.0064 ***	0.0054 ***	0.0012
	(0.0012)	(0.0031)	(0.0011)	(0.0048)	(0.0030)	(0.0026)	(0.0131)	(0.0124)	(0.0056)	(0.0030)	(0.0020)	(0.0017)	(0.0008)
ln(*PrivatefundingO*)$_{it}$	0.0032 ***	-0.0003	0.0029 ***	-0.0064	0.0027	0.0002	0.0005	-0.0240	-0.0076	-0.0005	-0.0008	-0.0011	0.0003
	(0.0010)	(0.0025)	(0.0009)	(0.0045)	(0.0032)	(0.0026)	(0.0180)	(0.0183)	(0.0059)	(0.0015)	(0.0013)	(0.0010)	(0.0006)
ln(*NFPfundingO*)$_{it}$	-0.0021 **	0.0013	-0.0016 *	0.0088 **	0.0026	0.0042 **	0.0020	-0.0169	0.0050	0.0000	0.0010	0.0008	-0.0009
	(0.0009)	(0.0021)	(0.0008)	(0.0041)	(0.0022)	(0.0020)	(0.0114)	(0.0149)	(0.0055)	(0.0017)	(0.0012)	(0.0010)	(0.0006)
Constant	0.2690 **	0.3782 *	0.2848 **	0.3647 *	0.2731 **	0.3033 **	0.7501	0.4237	0.3376 *	0.2333 **	0.2918 **	0.2705 **	0.3384 **
	(0.0233)	(0.0540)	(0.0216)	(0.0793)	(0.0418)	(0.0368)	(0.1817)	(0.1739)	(0.0990)	(0.0429)	(0.0289)	(0.0240)	(0.0134)
Number of observations	10124	1954	12078	1243	3265	4508	281	420	613	4466	8089	12555	35201
Number of scientists	1330	313	1643	336	708	1044	111	225	191	679	1131	1810	5634
χ2	. ***	537.41 ***	4479.74 ***	390.77 ***	923.51 ***	1258.02 ***	92.61 ***	144.61 ***	190.72 ***	880.63 ***	1841.27 ***	2734.85 ***	9506.55 ***
Average number of years	7.61	6.24	7.35	3.70	4.61	4.32	2.53	1.87	3.21	6.58	7.15	6.94	6.25
R^2 within groups	0.25	0.19	0.23	0.18	0.15	0.15	0.25	0.23	0.19	0.13	0.13	0.13	0.17
R^2 overall	0.32	0.25	0.31	0.28	0.26	0.26	0.31	0.28	0.28	0.21	0.24	0.23	0.27
R^2 between groups	0.49	0.26	0.44	0.33	0.38	0.35	0.27	0.25	0.27	0.28	0.41	0.37	0.35

*, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively. Year dummies and university dummies are significant. The definition of regression tags indicating the sample: A= Medical, B= Health Science, c= Business and Management, D= Social Science, E= Education, F= Humanities, G= Non-health professions, H= Engineering, I= Science.

APPENDIX G – VARIABLE DESCRIPTION OF CHAPTER 4[42]

Variable name	*Variable description*
$ColT90_{it}$ and $ColT95_{it}$	Dummy variables taking the value 1 if any of the coauthors of researcher *i* is amongst top funded scientists (top 10% and top 5% of total funding respectively)
$ColPub90_{it}$ and $ColPub95_{it}$	Dummy variables taking the value 1 if any of the coauthors of researcher *i* is amongst top funded scientists (top 10% and top 5% of public funding respectively)
$ColPriv90_{it}$, and $ColPriv95_{it}$	Dummy variables taking the value 1 if any of the coauthors of researcher *i* is amongst top funded scientists (top 10% and top 5% of private funding respectively)
$\ln(PublicfundingO_{it})$	Natural logarithm of the three-year average up to year *t* of public sector funding for the purpose of operational costs and direct expenditures of research of researcher *i*
$\ln(PrivatefundingO_{it})$	Natural logarithm of the three-year average up to year *t* of private sector funding for the purpose of operational costs and direct expenditures of research of researcher *i*
$\ln(NFPfundingO_{it})$	Natural logarithm of three-year average up to year *t* of funding from not-for-profit institutions (NFP) for the purpose of operational costs and direct expenditures of research of researcher *i*
$\ln(nbArticle_{it})$	Natural logarithm of the yearly number of articles published in year *t* by researcher *i*
$\ln(nbAuthor_{it})$	Natural logarithm of the three-year average of number of authors in the papers of researcher *i*
$PubORank_{it}$	Normalized rank of researcher *i* in the field in terms of three-year average of funding for the purpose of operational costs and direct expenditure of research
$nbScientistUni_{it}$	Number of scientists in the university and division[43] of researcher *i*
$\ln(totPublicfundingOcluster_{it})$	Natural logarithm of the three-year average up to year *t* of public sector funding for the purpose of operational costs and direct expenditures of research, which is aggregated over cluster[44] of researcher *i*
Age_{it}	Age of researcher *i* in year *t*
$dFemale_i$	Dummy variable taking the value 1 if researcher *i* is a woman and 0 otherwise
$dULaval_i$, $dUMcGill_i$, …, $dUdeM_i$	Dummy variables indicating the university affiliation of researcher *i*
$dMedical_i$, $dHumanities_i$, …, $dScience_i$	Dummy variables indicating the field of researcher *i*
d2000, d2001, …., d2012	Dummy variables indicating the year

[42] Some variables are transformed by natural logarithm function to be normal variables and satisfy the necessary conditions for the right hand side variables of the regression equations.

[43] There are 9 divisions of Basic Medical Sciences, Business & Management, Education, Engineering, Health Sciences, Humanities, Non-Health Professional, Sciences, and Social Sciences

[44] Cluster is more detail categorization of researchers. There are 42 clusters of Agricultural & Food Sciences, Anthropology, Archaeology & Sociology, Biology & Botany, Business, Chemical Engineering, Chemistry, Civil Engineering, Computer & Information Science, Dentistry, Earth & Ocean Sciences, Economics, Education, Electrical & Computer Engineering, Fine & Performing Arts, Foreign Languages Literature, Linguistic, French/English, General Medicine, Geography, History, Kinesiology / Physical Education, Laboratory Medicine, Law & Legal Studies, Library & Information Sciences, Mathematics, Mechanical & Industrial Engineering, Media & Communication Studies, Medical Specialties, Nursing, Other Engineering, Other Health Sciences, Other Social Sciences & Humanities, Philosophy, Physics & Astronomy, Planning & Architecture, Political Science, Psychology, Public Health & Health Administration, Rehabilitation Therapy, Religious Studies & Vocations, Resource Management & Forestry, Social Work, and Surgical Specialties.

APPENDIX H – SUMMARY STATISTICS OF CHAPTER 4

	Average	*Standard Deviation*	*Minimum*	*Maximum*
nbArticle	3.3880	3.6327	1	85
PublicfundingO	113,648	200,875	0	10,100,000
PrivatefundingO	21,609	96,627	0	5,508,461
NFPfundingO	21,197	126,788	0	8,404,625
dFemale	0.2348	0.4239	0	1
Age	50.482	9.209	16	92
nbAuthor	6.0821	34.6297	0.3333	2063.646
PubORank	0.6704	0.2314	0.0010	1
totPublicfundingOcluster	31,600,000	32,000,000	73615	121,000,000
nbScientistUni	452.84	446.59	5	1604
ColT90	0.4113	0.4921	0	1
ColT95	0.2766	0.4473	0	1
ColPub90	0.4113	0.4921	0	1
ColPub95	0.2712	0.4446	0	1
ColPriv90	0.3331	0.4713	0	1
ColPriv95	0.2165	0.4119	0	1

The variables are not summarized in logarithmic scale and they are raw amount. In some disciplines of Physics, there are many scientists involved in one project and therefore, the maximum for the number of authors is high. (No. observation: 32,299)

APPENDIX I – CORRELATION TABLE OF CHAPTER 4

		1	2	3	4	5	6	7	8	9	10
ln(nbArticle)	**1**	1									
ln(PublicfundingO)	**2**	0.2240	1								
ln(PrivatefundingO)	**3**	0.2205	0.0958	1							
ln(NFPfundingO)	**4**	0.2343	0.1647	0.2194	1						
dFemale	**5**	-0.1088	-0.0265	-0.131	-0.0217	1					
Age	**6**	0.0218	-0.0714	0.0072	-0.0462	-0.1185	1				
ln(nbAuthor)	**7**	0.4026	0.0578	0.185	0.2339	-0.0446	0.0772	1			
PubORank	**8**	0.2963	0.6635	0.1391	0.2050	-0.0349	0.0230	0.1483	1		
ln(totPublicfundingOcluster)	**9**	0.2078	0.0639	0.1278	0.2492	-0.0228	0.0425	0.3552	0.0438	1	
ln(nbScientistUni)	**10**	0.0711	-0.1094	0.0676	0.1021	-0.0239	0.0872	0.1974	0.0006	0.1830	1
ColT90	**11**	0.3788	0.1895	0.2532	0.2808	-0.0436	-0.0132	0.3616	0.2673	0.2508	0.1053
ColT95	**12**	0.3537	0.1586	0.2339	0.2494	-0.0411	-0.0082	0.3216	0.2213	0.2313	0.0942
ColPub90	**13**	0.3727	0.2261	0.1900	0.2266	-0.0305	-0.0143	0.3299	0.3084	0.2304	0.0891
ColPub95	**14**	0.3459	0.1922	0.1711	0.2028	-0.0322	-0.0158	0.2905	0.2536	0.2053	0.0752
ColPriv90	**15**	0.3133	0.0428	0.4560	0.2034	-0.0638	-0.0178	0.2874	0.0859	0.1990	0.1038
ColPriv95	**16**	0.2857	0.0349	0.3821	0.1847	-0.0502	-0.0236	0.2646	0.0690	0.1763	0.1046

		11	12	13	14	15	16
ColT90	**11**	1					
ColT95	**12**	0.7399	1				
ColPub90	**13**	0.7934	0.6522	1			
ColPub95	**14**	0.7299	0.7943	0.7299	1		
ColPriv90	**15**	0.5070	0.4803	0.3887	0.3801	1	
ColPriv95	**16**	0.4910	0.4996	0.3497	0.3660	0.7437	1

Significant at 1% level (No. observations = 32,299)

APPENDIX J – FIRST STAGE REGRESSION RESULTS OF TABLE 4.1

Dependent Var: PublicfundingO	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
dFemale$_i$	0.0676	0.0672	0.0674	0.0681	0.0644	0.0648
	0.0466	0.0464	0.0465	0.0461	0.0467	0.0466
Age$_{it}$	0.0272 **	0.0276 **	0.0260 **	0.0265 **	0.0290 **	0.0287 **
	0.0138	0.0138	0.0138	0.0137	0.0138	0.0138
Age$_{it}^2$	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***
	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
ln(nbAuthor$_{it}$)	-0.0659 **	-0.0609 **	-0.0687 **	-0.0661 **	-0.0449 *	-0.0474 *
	0.0267	0.0266	0.0266	0.0266	0.0266	0.0266
ln(PrivatefundingO$_{it}$)	0.0081 ***	0.0080 ***	0.0082 ***	0.0080 ***	0.0097 ***	0.0090 ***
	0.0028	0.0028	0.0028	0.0028	0.0029	0.0029
ln(NFPfundingO$_{it}$)	0.0150 ***	0.0155 ***	0.0154 ***	0.0155 ***	0.0160 ***	0.0159 ***
	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026
PubORank$_{it}$	7.4159 ***	7.4413 ***	7.4071 ***	7.4412 ***	7.4442 ***	7.4472 ***
	0.0612	0.0610	0.0613	0.0610	0.0610	0.0610
ln(totPublicfundingOcluster$_{it}$)	0.2871 ***	0.2885 ***	0.2840 ***	0.2866 ***	0.2918 ***	0.2915 ***
	0.0369	0.0367	0.0368	0.0365	0.0369	0.0368
ln(nbScientistUni$_{it}$)	-0.3863 ***	-0.3856 ***	-0.3865 ***	-0.3851 ***	-0.3855 ***	-0.3856 ***
	0.0298	0.0296	0.0297	0.0294	0.0298	0.0297
ColT90$_{it}$	0.1126 ***					
	0.0230					
ColT95$_{it}$		0.0813 ***				
		0.0247				
ColPub90$_{it}$			0.1347 ***			
			0.0226			
ColPub95$_{it}$				0.1183 ***		
				0.0243		
ColPriv90$_{it}$					-0.0349	
					0.0240	
ColPriv95$_{it}$						-0.0188
						0.0266
Constant	1.8282 ***	1.7941 ***	1.9162 ***	1.8485 ***	1.7043 **	1.7128 **
	0.8048	0.8017	0.8031	0.7978	0.8050	0.8038
Number of observations	32299	32299	32299	32299	32299	32299
$\chi 2$	18026	18096	18097	18223	17988	18017

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX K – FIRST STAGE REGRESSION RESULTS OF TABLE 4.2

Dependent Var: PublicfundingO	(1)	(2)	(3)	(4)	(5)	(6)
$dFemale_i$	0.0693	0.0675	0.0685	0.0682	0.0651	0.0656
	0.0467	0.0465	0.0465	0.0461	0.0467	0.0466
Age_{it}	0.0266 **	0.0278 **	0.0254 **	0.0265 **	0.0286 **	0.0282 **
	0.0138	0.0138	0.0138	0.0137	0.0138	0.0138
Age_{it}^2	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***
	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
$ln(nbAuthor_{it})$	-0.1447 ***	-0.0809 ***	-0.1530 ***	-0.0910 ***	-0.0753 ***	-0.0931 ***
	0.0314	0.0294	0.0308	0.0291	0.0297	0.0285
$ln(PrivatefundingO_{it})$	0.0082 ***	0.0081 ***	0.0083 ***	0.0081 ***	0.0099 ***	0.0093 ***
	0.0028	0.0028	0.0028	0.0028	0.0029	0.0029
$ln(NFPfundingO_{it})$	0.0151 ***	0.0155 ***	0.0155 ***	0.0156 ***	0.0160 ***	0.0160 ***
	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026
$PubORank_{it}$	7.4241 ***	7.4428 ***	7.4162 ***	7.4424 ***	7.4490 ***	7.4567 ***
	0.0612	0.0611	0.0613	0.0610	0.0610	0.0610
$ln(totPublicfundingOcluster_{it})$	0.2895 ***	0.2891 ***	0.2859 ***	0.2877 ***	0.2935 ***	0.2949 ***
	0.0369	0.0367	0.0368	0.0365	0.0369	0.0368
$ln(nbScientistUni_{it})$	-0.3844 ***	-0.3850 ***	-0.3845 ***	-0.3843 ***	-0.3846 ***	-0.3843 ***
	0.0298	0.0296	0.0296	0.0294	0.0298	0.0297
$ColT90_{it}$	-0.2042 ***					
	0.0704					
$ColT95_{it}$		-0.0428				
		0.0798				
$ColPub90_{it}$			-0.2173 ***			
			0.0687			
$ColPub95_{it}$				-0.0390		
				0.0778		
$ColPriv90_{it}$					-0.2040 ***	
					0.0778	
$ColPriv95_{it}$						-0.4041 ***
						0.0913
$ColT90_{it}*ln(nbArticle_{it})$	0.1898 ***					
	0.0399					
$ColT95_{it}*ln(nbArticle_{it})$		0.0705				
		0.0432				
$ColPub90_{it}*ln(nbArticle_{it})$			0.2106 ***			
			0.0388			
$ColPub95_{it}*ln(nbArticle_{it})$				0.0896 **		
				0.0421		
$ColPriv90_{it}*ln(nbArticle_{it})$					0.0996 **	
					0.0436	
$ColPriv95_{it}*ln(nbArticle_{it})$						0.2172 ***
						0.0492
Constant	1.8952 ***	1.8018 ***	1.9988 ***	1.8586 ***	1.7164 ***	1.7181 **
	0.8050	0.8024	0.8029	0.7983	0.8050	0.8038
Number of observations	32299	32299	32299	32299	32299	32299
$\chi 2$	18052	18079	18141	18215	17993	18040

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX L – FIRST STAGE REGRESSION RESULTS OF TABLE 4.3

Dependent Var: PublicfundingO	(1)	(2)	(3)	(4)	(5)	(6)
$dFemale_i$	0.0680	0.0744	0.0596	0.0771	0.0607	0.0543
	0.0495	0.0478	0.0493	0.0475	0.0485	0.0476
Age_{it}	0.0271 **	0.0275 **	0.0260 **	0.0263 **	0.0291 **	0.0289 **
	0.0138	0.0138	0.0138	0.0137	0.0138	0.0138
Age_{it}^2	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***
	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
$ln(nbAuthor_{it})$	-0.0663 ***	-0.0612 ***	-0.0692 ***	-0.0666 ***	-0.0450 ***	-0.0477 ***
	0.0267	0.0266	0.0266	0.0266	0.0266	0.0266
$ln(PrivatefundingO_{it})$	0.0080 ***	0.0080 ***	0.0082 ***	0.0079 ***	0.0097 ***	0.0091 ***
	0.0028	0.0028	0.0028	0.0028	0.0029	0.0029
$ln(NFPfundingO_{it})$	0.0150 ***	0.0155 ***	0.0154 ***	0.0155 ***	0.0160 ***	0.0159 ***
	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026
$PubORank_{it}$	7.4178 ***	7.4430 ***	7.4107 ***	7.4433 ***	7.4449 ***	7.4479 ***
	0.0612	0.0610	0.0613	0.0610	0.0610	0.0610
$ln(totPublicfundingOcluster_{it})$	0.2871 ***	0.2883 ***	0.2841 ***	0.2864 ***	0.2917 ***	0.2914 ***
	0.0368	0.0367	0.0367	0.0365	0.0369	0.0368
$ln(nbScientistUni_{it})$	-0.3862 ***	-0.3854 ***	-0.3865 ***	-0.3850 ***	-0.3855 ***	-0.3857 ***
	0.0297	0.0296	0.0296	0.0294	0.0298	0.0297
$ColT90_{it}$	0.1128 ***					
	0.0262					
$ColT95_{it}$		0.0893 ***				
		0.0280				
$ColPub90_{it}$			0.1289 ***			
			0.0257			
$ColPub95_{it}$				0.1282 ***		
				0.0276		
$ColPriv90_{it}$					-0.0384	
					0.0269	
$ColPriv95_{it}$						-0.0336
						0.0298
$ColT90_{it}*dFemale_i$	-0.0012					
	0.0520					
$ColT95_{it}*dFemale_i$		-0.0347				
		0.0571				
$ColPub90_{it}*dFemale_i$			0.0242			
			0.0511			
$ColPub95_{it}*dFemale_i$				-0.0429		
				0.0564		
$ColPriv90_{it}*dFemale_i$					0.0154	
					0.0548	
$ColPriv95_{it}*dFemale_i$						0.0684
						0.0626
Constant	1.8289 **	1.7960 **	1.9157 **	1.8533 **	1.7048 **	1.7128 **
	0.8042	0.8010	0.8021	0.7970	0.8048	0.8037
Number of observations	32299	32299	32299	32299	32299	32299
$\chi 2$	18042	18115	18125	18248	17991	18022

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX M – FIRST STAGE REGRESSION RESULTS OF TABLE 4.4

Dependent Var: PublicfundingO	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
dFemale$_i$	0.0678	0.0674	0.0678	0.0684	0.0648	0.0651
	0.0464	0.0463	0.0461	0.0459	0.0467	0.0466
Age$_{it}$	0.0270 **	0.0273 **	0.0255 **	0.0261 **	0.0290 **	0.0287 **
	0.0138	0.0138	0.0137	0.0137	0.0138	0.0138
Age$_{it}^2$	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***
	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
ln(nbAuthor$_{it}$***)***	-0.0665 ***	-0.0609 ***	-0.0702 ***	-0.0669 ***	-0.0434	-0.0464 *
	0.0267	0.0266	0.0266	0.0265	0.0266	0.0266
ln(PrivatefundingO$_{it}$***)***	0.0049	0.0059 *	0.0054	0.0067 **	0.0057	0.0072 **
	0.0036	0.0033	0.0035	0.0032	0.0036	0.0032
ln(NFPfundingO$_{it}$***)***	0.0150 ***	0.0154 ***	0.0154 ***	0.0155 ***	0.0159 ***	0.0159 ***
	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026
PubORank$_{it}$	7.4241 ***	7.4469 ***	7.4215 ***	7.4489 ***	7.4445 ***	7.4471 ***
	0.0612	0.0610	0.0612	0.0610	0.0610	0.0610
ln(totPublicfundingOcluster$_{it}$***)***	0.2876 ***	0.2890 ***	0.2846 ***	0.2868 ***	0.2923 ***	0.2918 ***
	0.0367	0.0366	0.0365	0.0364	0.0369	0.0368
ln(nbScientistUni$_{it}$***)***	-0.3853 ***	-0.3846 ***	-0.3854 ***	-0.3845 ***	-0.3850 ***	-0.3853 ***
	0.0296	0.0295	0.0294	0.0293	0.0298	0.0297
ColT90$_{it}$	0.0914 ***					
	0.0280					
ColT95$_{it}$		0.0569 *				
		0.0312				
ColPub90$_{it}$			0.1155 ***			
			0.0276			
ColPub95$_{it}$				0.1048 ***		
				0.0305		
ColPriv90$_{it}$					-0.0709 **	
					0.0309	
ColPriv95$_{it}$						-0.0518
						0.0369
ColT90$_{it}$****ln(PrivatefundingO***$_{it}$***)***	0.0059					
	0.0044					
ColT95$_{it}$****ln(PrivatefundingO***$_{it}$***)***		0.0058				
		0.0045				
ColPub90$_{it}$****ln(PrivatefundingO***$_{it}$***)***			0.0052			
			0.0043			
ColPub95$_{it}$****ln(PrivatefundingO***$_{it}$***)***				0.0033		
				0.0045		
ColPriv90$_{it}$****ln(PrivatefundingO***$_{it}$***)***					0.0085 *	
					0.0046	
ColPriv95$_{it}$****ln(PrivatefundingO***$_{it}$***)***						0.0062
						0.0048
Constant	1.8231 **	1.7878 **	1.9123 **	1.8495 **	1.6995 **	1.7087 **
	0.8022	0.8000	0.7984	0.7954	0.8049	0.8039
Number of observations	32299	32299	32299	32299	32299	32299
$\chi 2$	18102	18146	18232	18293	17993	18017

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX N – FIRST STAGE REGRESSION RESULTS OF TABLE 4.5

Dependent Var: PublicfundingO	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
dFemale$_{i}$	0.0683	0.0678	0.0681	0.0687	0.0645	0.0647
	0.0464	0.0463	0.0462	0.0459	0.0465	0.0464
Age$_{it}$	0.0268 **	0.0271 **	0.0256 *	0.0260 *	0.0288 **	0.0287 **
	0.0138	0.0137	0.0137	0.0137	0.0138	0.0138
Age$_{it}^{2}$	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***	-0.0005 ***
	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
ln(nbAuthor$_{it}$***)***	-0.0657 **	-0.0601 **	-0.0700 ***	-0.0665 **	-0.0457 *	-0.0474 *
	0.0267	0.0266	0.0266	0.0265	0.0266	0.0266
ln(PrivatefundingO$_{it}$***)***	0.0077 ***	0.0078 ***	0.0079 ***	0.0077 ***	0.0095 ***	0.0089 ***
	0.0028	0.0028	0.0028	0.0028	0.0029	0.0029
ln(NFPfundingO$_{it}$***)***	0.0096 ***	0.0122 ***	0.0127 ***	0.0130 ***	0.0139 ***	0.0137 ***
	0.0034	0.0030	0.0033	0.0030	0.0031	0.0029
PubORank$_{it}$	7.4242 ***	7.4474 ***	7.4211 ***	7.4520 ***	7.4535 ***	7.4538 ***
	0.0612	0.0610	0.0612	0.0610	0.0609	0.0609
ln(totPublicfundingOcluster$_{it}$***)***	0.2875 ***	0.2888 ***	0.2843 ***	0.2871 ***	0.2921 ***	0.2918 ***
	0.0367	0.0366	0.0365	0.0363	0.0367	0.0367
ln(nbScientistUni$_{it}$***)***	-0.3850 ***	-0.3848 ***	-0.3858 ***	-0.3843 ***	-0.3849 ***	-0.3847 ***
	0.0296	0.0295	0.0294	0.0293	0.0296	0.0296
ColT90$_{it}$	0.0681 **					
	0.0289					
ColT95$_{it}$		0.0354				
		0.0326				
ColPub90$_{it}$			0.1114 ***			
			0.0285			
ColPub95$_{it}$				0.0807 **		
				0.0321		
ColPriv90$_{it}$					-0.0601 **	
					0.0306	
ColPriv95$_{it}$						-0.0626 *
						0.0355
ColT90$_{it}$****ln(NFPfundingO***$_{it}$***)***	0.0109 **					
	0.0043					
ColT95$_{it}$****ln(NFPfundingO***$_{it}$***)***		0.0097 **				
		0.0045				
ColPub90$_{it}$****ln(NFPfundingO***$_{it}$***)***			0.0056			
			0.0042			
ColPub95$_{it}$****ln(NFPfundingO***$_{it}$***)***				0.0080 *		
				0.0044		
ColPriv90$_{it}$****ln(NFPfundingO***$_{it}$***)***					0.0055	
					0.0043	
ColPriv95$_{it}$****ln(NFPfundingO***$_{it}$***)***						0.0088 *
						0.0048
Constant	1.8350 **	1.8005 **	1.9188 ***	1.8483 **	1.7038 **	1.7083 **
	0.8021	0.7997	0.7988	0.7947	0.8022	0.8019
Number of observations	32299	32299	32299	32299	32299	32299
$\chi 2$	18109	18158	18219	18317	18066	18075

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX O – OLS REGRESSION RESULTS OF TABLE 4.1 (CLUSTERED GROUPS – NO ENDOGENEITY ASSUMPTION)

Dependent Var: nbArticle	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
dFemale$_i$	-0.0933 ***	-0.0923 ***	-0.0947 ***	-0.0931 ***	-0.0947 ***	-0.0941 ***
	(0.0063)	(0.0063)	(0.0063)	(0.0063)	(0.0063)	(0.0064)
Age$_{it}$	0.0163 ***	0.0166 ***	0.0156 ***	0.0163 ***	0.0176 ***	0.0173 ***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0025)
Age$_{it}^2$	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
ln(*nbAuthor$_{it}$*)	0.2444 ***	0.2514 ***	0.2466 ***	0.2533 ***	0.2644 ***	0.2683 ***
	(0.0054)	(0.0053)	(0.0053)	(0.0053)	(0.0053)	(0.0053)
ln(*PublicfundingO$_{it}$*)	0.0289 ***	0.0303 ***	0.0267 ***	0.0285 ***	0.0352 ***	0.0351 ***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
ln(*PrivatefundingO$_{it}$*)	0.0103 ***	0.0106 ***	0.0117 ***	0.0119 ***	0.0053 ***	0.0076 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
ln(*NFPfundingO$_{it}$*)	0.0090 ***	0.0096 ***	0.0102 ***	0.0105 ***	0.0113 ***	0.0114 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
ColT90$_{it}$	0.2462 ***					
	(0.0060)					
ColT95$_{it}$		0.2566 ***				
		(0.0064)				
ColPub90$_{it}$			0.2520 ***			
			(0.0058)			
ColPub95$_{it}$				0.2626 ***		
				(0.0063)		
ColPriv90$_{it}$					0.2239 ***	
					(0.0065)	
ColPriv95$_{it}$						0.2241 ***
						(0.0071)
Constant	0.0985	0.0999	0.1309	0.1157 *	0.0012	0.0217
	(0.0630)	(0.0631)	(0.0629)	(0.0630)	(0.0634)	(0.0636)
Number of observations	35684	35684	35684	35684	35684	35684
Log likelihood	-30643	-30643	-30643	-30643	-30643	-30643
Aikaike information criterion	49475	49574	49299	49445	49963	50154
Bayesian information criterion	49755	49854	49579	49725	50243	50434
R^2	0.2831	0.2811	0.2866	0.2837	0.2732	0.2693
Adjusted R^2	0.2825	0.2805	0.2860	0.2831	0.2726	0.2687
F-test	439	435	447	441	418	410

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX P – OLS REGRESSION RESULTS OF TABLE 4.2 (CLUSTERED GROUPS – NO ENDOGENEITY ASSUMPTION)

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
$dFemale_i$	-0.0934 ***	-0.0924 ***	-0.0945 ***	-0.0930 ***	-0.0947 ***	-0.0944 ***
	(0.0063)	(0.0063)	(0.0063)	(0.0063)	(0.0063)	(0.0064)
Age_{it}	0.0163 ***	0.0165 ***	0.0156 ***	0.0164 ***	0.0176 ***	0.0173 ***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0025)
Age_{it}^2	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\ln(nbAuthor_{it})$	0.2468 ***	0.2599 ***	0.2347 ***	0.2486 ***	0.2643 ***	0.2786 ***
	(0.0067)	(0.0061)	(0.0066)	(0.0061)	(0.00610	(0.0059)
$\ln(PublicfundingO_{it})$	0.0289 ***	0.0302 ***	0.0266 ***	0.0285 ***	0.0352 ***	0.0350 ***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
$\ln(PrivatefundingO_{it})$	0.0103 ***	0.0106 ***	0.0117 ***	0.0119 ***	0.0053 ***	0.0076 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
$\ln(NFPfundingO_{it})$	0.0090 ***	0.0096 ***	0.0102 ***	0.0105 ***	0.0113 ***	0.0114 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
$ColT90_{it}$	0.2559 ***					
	(0.0179)					
$ColT95_{it}$		0.3123 ***				
		(0.0208)				
$ColPub90_{it}$			0.2016 ***			
			(0.0176)			
$ColPub95_{it}$				0.2310 ***		
				(0.0205)		
$ColPriv90_{it}$					0.2231 ***	
					(0.0197)	
$ColPriv95_{it}$						0.3182 ***
						(0.0232)
$ColT90_{it}*\ln(nbAuthor_{it})$	-0.0058 ***					
	(0.0101)					
$ColT95_{it}*\ln(nbAuthor_{it})$		-0.0313 ***				
		(0.0111)				
$ColPub90_{it}*\ln(nbAuthor_{it})$			0.0302 ***			
			(0.0099)			
$ColPub95_{it}*\ln(nbAuthor_{it})$				0.0179 ***		
				(0.0111)		
$ColPriv90_{it}*\ln(nbAuthor_{it})$					0.0004 ***	
					(0.0108)	
$ColPriv95_{it}*\ln(nbAuthor_{it})$						-0.0522 ***
						(0.0123)
Constant	0.0952	0.0899	0.1479 **	0.1219 **	0.0014	0.0099
	(0.0633)	(0.0632)	(0.0631)	(0.0631)	(0.0635)	(0.0636)
Number of observations	35684	35684	35684	35684	35684	35684
Log likelihood	-30643	-30643	-30643	-30643	-30643	-30643
Aikaike information criterion	49477	49568	49292	49445	49965	50137
Bayesian information criterion	49765	49857	49581	49733	50254	50426
R^2	0.2831	0.2813	0.2868	0.2838	0.2732	0.2697
Adjusted R^2	0.2824	0.2806	0.2861	0.2831	0.2726	0.2690
F-test	426	422	434	427	406	398

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX Q – OLS REGRESSION RESULTS OF TABLE 4.3 (CLUSTERED GROUPS – NO ENDOGENEITY ASSUMPTION)

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
$dFemale_i$	-0.0800 ***	-0.0807 ***	-0.0700 ***	-0.0794 ***	-0.0887 ***	-0.0901 ***
	(0.0079)	(0.0072)	(0.0079)	(0.0072)	(0.0075)	(0.0070)
Age_{it}	0.0162 ***	0.0166 ***	0.0155 ***	0.0162 ***	0.0176 ***	0.0173 ***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0025)
Age_{it}^2	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\ln(nbAuthor_{it})$	0.2444 ***	0.2515 ***	0.2467 ***	0.2534 ***	0.2645 ***	0.2683 ***
	(0.0054)	(0.0053)	(0.0053)	(0.0053)	(0.0053)	(0.0053)
$\ln(PublicfundingO_{it})$	0.0289 ***	0.0302 ***	0.0266 ***	0.0285 ***	0.0352 ***	0.0351 ***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
$\ln(PrivatefundingO_{it})$	0.0102 ***	0.0105 ***	0.0116 ***	0.0119 ***	0.0052 ***	0.0075 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
$\ln(NFPfundingO_{it})$	0.0090 ***	0.0096 ***	0.0101 ***	0.0105 ***	0.0113 ***	0.0114 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
$ColT90_{it}$	0.2542 ***					
	(0.0066)					
$ColT95_{it}$		0.2668 ***				
		(0.0071)				
$ColPub90_{it}$			0.2667 ***			
			(0.0065)			
$ColPub95_{it}$				0.2748 ***		
				(0.0070)		
$ColPriv90_{it}$					0.2284 ***	
					(0.0071)	
$ColPriv95_{it}$						0.2286 ***
						(0.0078)
$ColT90_{it}*dFemale_i$	-0.0349 ***					
	(0.0125)					
$ColT95_{it}*dFemale_i$		-0.0461 ***				
		(0.0141)				
$ColPub90_{it}*dFemale_i$			-0.0636 ***			
			(0.0125)			
$ColPub95_{it}*dFemale_i$				-0.0546 ***		
				(0.0140)		
$ColPriv90_{it}*dFemale_i$					-0.0202 ***	
					(0.0134)	
$ColPriv95_{it}*dFemale_i$						-0.0211 ***
						(0.0156)
Constant	0.0978	0.0986	0.1282 **	0.1146 *	-0.0001	0.0209
	(0.0630)	(0.0631)	(0.0629)	(0.0630)	(0.0634)	(0.0636)
Number of observations	35684	35684	35684	35684	35684	35684
Log likelihood	-30643	-30643	-30643	-30643	-30643	-30643
Aikaike information criterion	49470	49566	49275	49432	49963	50154
Bayesian information criterion	49758	49854	49564	49721	50251	50442
R^2	0.2833	0.2813	0.2871	0.2840	0.2733	0.2694
Adjusted R^2	0.2826	0.2807	0.2865	0.2833	0.2726	0.2687
F-test	426	422	435	428	406	398

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX R – OLS REGRESSION RESULTS OF TABLE 4.4 (CLUSTERED GROUPS – NO ENDOGENEITY ASSUMPTION)

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
$dFemale_i$	-0.0927***	-0.0919***	-0.0942***	-0.0928***	-0.0948***	-0.0942***
	(0.0063)	(0.0063)	(0.0063)	(0.0063)	(0.0063)	(0.0064)
Age_{it}	0.0166***	0.0168***	0.0159***	0.0164***	0.0176***	0.0173***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0025)
Age_{it}^2	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\ln(nbAuthor_{it})$	0.2455***	0.2523***	0.2488***	0.2544***	0.2642***	0.2682***
	(0.0054)	(0.0053)	(0.0053)	(0.0053)	(0.0053)	(0.0053)
$\ln(PublicfundingO_{it})$	0.0286***	0.0301***	0.0262***	0.0283***	0.0352***	0.0351***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
$\ln(PrivatefundingO_{it})$	0.0052***	0.0080***	0.0055***	0.0091***	0.0058***	0.0077***
	(0.0009)	(0.0007)	(0.0008)	(0.0007)	(0.0009)	(0.0008)
$\ln(NFPfundingO_{it})$	0.0089***	0.0095***	0.0100***	0.0104***	0.0113***	0.0114***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
$ColT90_{it}$	0.2131***					
	(0.0072)					9
$ColT95_{it}$		0.2276***				
		(0.0081)				
$ColPub90_{it}$			0.2077***			
			(0.0070)			
$ColPub95_{it}$				0.2286***		
				(0.0079)		
$ColPriv90_{it}$					0.2279***	
					(0.0082)	
$ColPriv95_{it}$						0.2253***
						(0.0098)
$ColT90_{it}*\ln(PrivatefundingO_{it})$	0.0094***					
	(0.0011)					
$ColT95_{it}*\ln(PrivatefundingO_{it})$		0.0068***				
		(0.0012)				
$ColPub90_{it}*\ln(PrivatefundingO_{it})$			0.0124***			
			(0.0011)			
$ColPub95_{it}*\ln(PrivatefundingO_{it})$				0.0084***		
				(0.0012)		
$ColPriv90_{it}*\ln(PrivatefundingO_{it})$					-0.0010***	
					(0.0012)	
$ColPriv95_{it}*\ln(PrivatefundingO_{it})$						-0.0002***
						(0.0013)
Constant	0.1051*	0.1029	0.1423**	0.1223*	0.0008	0.0216
	(0.0630)	(0.0631)	(0.0628)	(0.0630)	(0.0634)	(0.0636)
Number of observations	35684	35684	35684	35684	35684	35684
Log likelihood	-30643	-30643	-30643	-30643	-30643	-30643
Aikaike information criterion	49409	49542	49177	49397	49965	50156
Bayesian information criterion	49697	49831	49466	49686	50253	50444
R^2	0.2845	0.2818	0.2891	0.2847	0.2732	0.2693
Adjusted R^2	0.2838	0.2811	0.2884	0.2840	0.2726	0.2687
F-test	429	423	439	429	406	398

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX S – OLS REGRESSION RESULTS OF TABLE 4.5 (CLUSTERED GROUPS – NO ENDOGENEITY ASSUMPTION)

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
dFemale$_i$	-0.0925 ***	-0.0918 ***	-0.0940 ***	-0.0927 ***	-0.0950 ***	-0.0943 ***
	(0.0063)	(0.0063)	(0.0063)	(0.0063)	(0.0063)	(0.0064)
Age$_{it}$	0.0164 ***	0.0166 ***	0.0157 ***	0.0163 ***	0.0177 ***	0.0175 ***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0025)
Age$_{it}^2$	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
ln(*nbAuthor$_{it}$*)	0.2455 ***	0.2526 ***	0.2481 ***	0.2547 ***	0.2651 ***	0.2692 ***
	(0.0054)	(0.0053)	(0.0053)	(0.0053)	(0.0053)	(0.0053)
ln(*PublicfundingO$_{it}$*)	0.0286 ***	0.0300 ***	0.0264 ***	0.0284 ***	0.0349 ***	0.0349 ***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
ln(*PrivatefundingO$_{it}$*)	0.0102 ***	0.0105 ***	0.0116 ***	0.0118 ***	0.0051 ***	0.0075 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
ln(*NFPfundingO$_{it}$*)	0.0046 ***	0.0069 ***	0.0053 ***	0.0076 ***	0.0074 ***	0.0089 ***
	(0.0008)	(0.0007)	(0.0008)	(0.0007)	(0.0007)	(0.0007)
ColT90$_{it}$	0.2113 ***					
	(0.0074)					
ColT95$_{it}$		0.2187 ***				
		(0.0085)				
ColPub90$_{it}$			0.2119 ***			
			(0.0073)			
ColPub95$_{it}$				0.2220 ***		
				(0.0083)		
ColPriv90$_{it}$					0.1815 ***	
					(0.0081)	
ColPriv95$_{it}$						0.1772 ***
						(0.0094)
ColT90$_{it}$*ln(NFPfundingO$_{it}$)	0.0088 ***					
	(0.0011)					
ColT95$_{it}$*ln(NFPfundingO$_{it}$)		0.0079 ***				
		(0.0012)				
ColPub90$_{it}$*ln(NFPfundingO$_{it}$)			0.0100 ***			
			(0.0011)			
ColPub95$_{it}$*ln(NFPfundingO$_{it}$)				0.0088 ***		
				(0.0012)		
ColPriv90$_{it}$*ln(NFPfundingO$_{it}$)					0.0099 ***	
					(0.0011)	
ColPriv95$_{it}$*ln(NFPfundingO$_{it}$)						0.0097 ***
						(0.0013)
Constant	0.1116 *	0.1100 *	0.1433 **	0.1246 *	0.0146	0.0289
	(0.0630)	(0.0631)	(0.0628)	(0.0630)	(0.0633)	(0.0635)
Number of observations	35684	35684	35684	35684	35684	35684
Log likelihood	-30643	-30643	-30643	-30643	-30643	-30643
Aikaike information criterion	49416	49531	49219	49392	49889	50097
Bayesian information criterion	49705	49819	49507	49680	50177	50386
R^2	0.2843	0.2820	0.2883	0.2848	0.2748	0.2705
Adjusted R^2	0.2837	0.2814	0.2876	0.2842	0.2741	0.2699
F-test	429	424	437	430	409	400

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX T – PANEL REGRESSION RESULTS WITHOUT ENDOGENEITY ASSUMPTION FOR TABLE 4.1

Dependent Var: nbArticle	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
dFemale$_i$	-0.0868 ***	-0.0865 ***	-0.0878 ***	-0.0876 ***	-0.0872 ***	-0.0869 ***
	(0.0091)	(0.0091)	(0.0091)	(0.0090)	(0.0092)	(0.0092)
Age$_{it}$	0.0353 ***	0.0358 ***	0.0339 ***	0.0348 ***	0.0375 ***	0.0377 ***
	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)
Age$_{it}^2$	-0.0004 ***	-0.0004 ***	-0.0003 ***	-0.0004 ***	-0.0004 ***	-0.0004 ***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
ln(*nbAuthor$_{it}$*)	0.1836 ***	0.1903 ***	0.1869 ***	0.1943 ***	0.1977 ***	0.2012 ***
	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0059)
ln(*PublicfundingO$_{it}$*)	0.0205 ***	0.0217 ***	0.0197 ***	0.0212 ***	0.0239 ***	0.0239 ***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
ln(*PrivatefundingO$_{it}$*)	0.0077 ***	0.0078 ***	0.0085 ***	0.0087 ***	0.0041 ***	0.0060 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0007)
ln(*NFPfundingO$_{it}$*)	0.0069 ***	0.0075 ***	0.0077 ***	0.0080 ***	0.0083 ***	0.0085 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
ColT90$_{it}$	0.1987 ***					
	(0.0055)					
ColT95$_{it}$		0.1998 ***				
		(0.0059)				
ColPub90$_{it}$			0.1982 ***			
			(0.0054)			
ColPub95$_{it}$				0.1966 ***		
				(0.0058)		
ColPriv90$_{it}$					0.1777 ***	
					(0.0058)	
ColPriv95$_{it}$						0.1719 ***
						(0.0064)
Constant	-0.1566 **	-0.1634 **	-0.1164	-0.1403 **	-0.2485 ***	-0.2428 ***
	(0.0732)	(0.0732)	(0.0730)	(0.0730)	(0.0736)	(0.0738)
Number of observations	35684	35684	35684	35684	35684	35684
Number of scientists	5772	5772	5772	5772	5772	5772
$\chi 2$	7451.4	7301.32	7575.32	7377.75	6991.21	6740.93
sigma	0.3735	0.3744	0.3736	0.3747	0.3746	0.3757
rho	0.2503	0.2463	0.2461	0.2407	0.2527	0.2517
R^2 within groups	0.0636	0.0589	0.0635	0.0573	0.0571	0.0514
R^2 overall	0.2769	0.2746	0.2806	0.2772	0.2662	0.2621
R^2 between groups	0.4640	0.4647	0.4698	0.4702	0.4544	0.4514

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX U – PANEL REGRESSION RESULTS WITHOUT ENDOGENEITY ASSUMPTION FOR TABLE 4.2

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
dFemale$_i$	-0.0867 ***	-0.0866 ***	-0.0877 ***	-0.0876 ***	-0.0871 ***	-0.0870 ***
	(0.0091)	(0.0091)	(0.0090)	(0.0090)	(0.0092)	(0.0092)
Age$_{it}$	0.0352 ***	0.0357 ***	0.0335 ***	0.0347 ***	0.0374 ***	0.0377 ***
	(0.0029)	(0.0029)	(0.0028)	(0.0028)	(0.0029)	(0.0029)
Age$_{it}^2$	-0.0004 ***	-0.0004 ***	-0.0003 ***	-0.0003 ***	-0.0004 ***	-0.0004 ***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
ln(*nbAuthor$_{it}$*)	0.1783 ***	0.1924 ***	0.1727 ***	0.1877 ***	0.1908 ***	0.2056 ***
	(0.0070)	(0.0066)	(0.0068)	(0.0064)	(0.0067)	(0.0064)
ln(*PublicfundingO$_{it}$*)	0.0206 ***	0.0217 ***	0.0197 ***	0.0212 ***	0.0239 ***	0.0238 ***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
ln(*PrivatefundingO$_{it}$*)	0.0077 ***	0.0078 ***	0.0086 ***	0.0087 ***	0.0042 ***	0.0060 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0007)
ln(*NFPfundingO$_{it}$*)	0.0069 ***	0.0075 ***	0.0078 ***	0.0080 ***	0.0083 ***	0.0085 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
ColT90$_{it}$	0.1762 ***					
	(0.0166)					
ColT95$_{it}$		0.2109 ***				
		(0.0190)				
ColPub90$_{it}$			0.1324 ***			
			(0.0164)			
ColPub95$_{it}$				0.1486 ***		
				(0.0188)		
ColPriv90$_{it}$					0.1397 ***	
					(0.0183)	
ColPriv95$_{it}$						0.2081 ***
						(0.0214)
ColT90$_{it}$*ln(nbAuthor$_{it}$)	0.0135					
	(0.0094)					
ColT95$_{it}$*ln(nbAuthor$_{it}$)		-0.0062				
		(0.0102)				
ColPub90$_{it}$*ln(nbAuthor$_{it}$)			0.0396 ***			
			(0.0092)			
ColPub95$_{it}$*ln(nbAuthor$_{it}$)				0.0275 ***		
				(0.0102)		
ColPriv90$_{it}$*ln(nbAuthor$_{it}$)					0.0224 **	
					(0.0102)	
ColPriv95$_{it}$*ln(nbAuthor$_{it}$)						-0.0203 *
						(0.0115)
Constant	-0.1467 **	-0.1643 **	-0.0886	-0.1267 *	-0.2378 ***	-0.2489 ***
	(0.0733)	(0.0732)	(0.0730)	(0.0730)	(0.0737)	(0.0738)
Number of observations	35684	35684	35684	35684	35684	35684
Number of scientists	5772	5772	5772	5772	5772	5772
$\chi 2$	7474.97	7322.29	7670.03	7434.88	7001.65	6743.49
sigma	0.3734	0.3743	0.3734	0.3746	0.3745	0.3757
rho	0.2488	0.2449	0.2412	0.2375	0.2523	0.2518
R^2 within groups	0.0638	0.0589	0.0639	0.0575	0.0574	0.0513
R^2 overall	0.2769	0.2747	0.2809	0.2773	0.2660	0.2624
R^2 between groups	0.4643	0.4648	0.4720	0.4712	0.4545	0.4517

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX V – PANEL REGRESSION RESULTS WITHOUT ENDOGENEITY ASSUMPTION FOR TABLE 4.3

Dependent Var: nbArticle	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
dFemale$_i$	-0.0896 ***	-0.0887 ***	-0.0799 ***	-0.0878 ***	-0.0883 ***	-0.0908 ***
	(0.0099)	(0.0095)	(0.0099)	(0.0094)	(0.0098)	(0.0095)
Age$_{it}$	0.0352 ***	0.0358 ***	0.0337 ***	0.0348 ***	0.0375 ***	0.0377 ***
	(0.0029)	(0.0029)	(0.0028)	(0.0028)	(0.0029)	(0.0029)
Age$_{it}^2$	-0.0004 ***	-0.0004 ***	-0.0003 ***	-0.0003 ***	-0.0004 ***	-0.0004 ***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
ln(*nbAuthor$_{it}$*)	0.1839 ***	0.1905 ***	0.1873 ***	0.1946 ***	0.1978 ***	0.2012 ***
	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0059)
ln(*PublicfundingO$_{it}$*)	0.0206 ***	0.0217 ***	0.0197 ***	0.0212 ***	0.0239 ***	0.0239 ***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
ln(*PrivatefundingO$_{it}$*)	0.0077 ***	0.0078 ***	0.0085 ***	0.0087 ***	0.0041 ***	0.0061 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0007)
ln(*NFPfundingO$_{it}$*)	0.0069 ***	0.0075 ***	0.0077 ***	0.0080 ***	0.0083 ***	0.0085 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
ColT90$_{it}$	0.1967 ***					
	(0.0062)					
ColT95$_{it}$		0.1976 ***				
		(0.0067)				
ColPub90$_{it}$			0.2040 ***			
			(0.00610			
ColPub95$_{it}$				0.1966 ***		
				(0.0066)		
ColPriv90$_{it}$					0.1769 ***	
					(0.0065)	
ColPriv95$_{it}$						0.1668 ***
						(0.0072)
ColT90$_{it}$*dFemale$_i$	0.0087					
	(0.0122)					
ColT95$_{it}$*dFemale$_i$		0.0101				
		(0.0136)				
ColPub90$_{it}$*dFemale$_i$			-0.0239 **			
			(0.0120)			
ColPub95$_{it}$*dFemale$_i$				0.0008		
				(0.0134)		
ColPriv90$_{it}$*dFemale$_i$					0.0040	
					(0.0130)	
ColPriv95$_{it}$*dFemale$_i$						0.0238
						(0.0149)
Constant	-0.1556 **	-0.1625 **	-0.1156	-0.1391 **	-0.2480 ***	-0.2426 ***
	(0.0731)	(0.0731)	(0.0729)	(0.0729)	(0.0736)	(0.0737)
Number of observations	35684	35684	35684	35684	35684	35684
Number of scientists	5772	5772	5772	5772	5772	5772
$\chi 2$	7467.68	7314.53	7602.34	7397.69	6998.72	6751.82
sigma	0.3734	0.3743	0.3736	0.3747	0.3746	0.3757
rho	0.2492	0.2454	0.2445	0.2393	0.2521	0.2512
R^2 within groups	0.0637	0.0590	0.0633	0.0573	0.0572	0.0516
R^2 overall	0.2769	0.2745	0.2810	0.2772	0.2662	0.2619
R^2 between groups	0.4638	0.4645	0.4708	0.4703	0.4543	0.4509

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX W – PANEL REGRESSION RESULTS WITHOUT ENDOGENEITY ASSUMPTION FOR TABLE 4.4

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
***dFemale**$_i$*	-0.0867***	-0.0865***	-0.0879***	-0.0877***	-0.0875***	-0.0870***
	(0.0091)	(0.0090)	(0.0090)	(0.0090)	(0.0092)	(0.0092)
***Age**$_{it}$*	0.0351***	0.0355***	0.0334***	0.0344***	0.0375***	0.0377***
	(0.0029)	(0.0029)	(0.0028)	(0.0028)	(0.0029)	(0.0029)
***Age**$_{it}^2$*	-0.0004***	-0.0004***	-0.0003***	-0.0003***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
ln(*nbAuthor*$_{it}$)	0.1852***	0.1916***	0.1896***	0.1960***	0.1970***	0.2010***
	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0059)
ln(*PublicfundingO*$_{it}$)	0.0206***	0.0218***	0.0197***	0.0212***	0.0239***	0.0239***
	(0.0010)	(0.0010)	(0.0010)	(0.00100	(0.0010)	(0.0010)
ln(*PrivatefundingO*$_{it}$)	0.0056***	0.0071***	0.0056***	0.0078***	0.0059***	0.0064***
	(0.0008)	(0.0008)	(0.0008)	(0.0007)	(0.0009)	(0.0008)
ln(*NFPfundingO*$_{it}$)	0.0069***	0.0075***	0.0077***	0.0080***	0.0083***	0.0085***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
***ColT90**$_{it}$*	0.1843***					
	(0.0066)					
***ColT95**$_{it}$*		0.1915***				
		(0.0074)				
***ColPub90**$_{it}$*			0.1768***			
			(0.0065)			
***ColPub95**$_{it}$*				0.1856***		
				(0.0073)		
***ColPriv90**$_{it}$*					0.1921***	
					(0.0074)	
***ColPriv95**$_{it}$*						0.1780***
						(0.0088)
ColT90**$_{it}$ln(PrivatefundingO**$_{it}$**)***	0.0042***					
	(0.0010)					
ColT95**$_{it}$ln(PrivatefundingO**$_{it}$**)***		0.0021**				
		(0.0011)				
ColPub90**$_{it}$ln(PrivatefundingO**$_{it}$**)***			0.0062***			
			(0.0010)			
ColPub95**$_{it}$ln(PrivatefundingO**$_{it}$**)***				0.0029***		
				(0.0011)		
ColPriv90**$_{it}$ln(PrivatefundingO**$_{it}$**)***					-0.0035***	
					(0.0011)	
ColPriv95**$_{it}$ln(PrivatefundingO**$_{it}$**)***						-0.0012
						(0.0012)
Constant	-0.1490**	-0.1577**	-0.1030	-0.1317*	-0.2499***	-0.2431***
	(0.0730)	(0.0730)	(0.0726)	(0.0728)	(0.0736)	(0.0738)
(Number of observations)	35684	35684	35684	35684	35684	35684
Number of scientists	5772	5772	5772	5772	5772	5772
$\chi 2$	7539.92	7371.85	7742.08	7475.38	7006.08	6744.94
sigma	0.3735	0.3744	0.3735	0.3747	0.3745	0.3757
rho	0.2455	0.2418	0.2379	0.2348	0.2524	0.2515
R^2 within groups	0.0634	0.0587	0.0634	0.0570	0.0576	0.0515
R^2 overall	0.2782	0.2752	0.2830	0.2781	0.2660	0.2621
R^2 between groups	0.4666	0.4662	0.4743	0.4722	0.4537	0.4513

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX X – PANEL REGRESSION RESULTS WITHOUT ENDOGENEITY ASSUMPTION FOR TABLE 4.5

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
dFemale$_i$	-0.0866 ***	-0.0864 ***	-0.0877 ***	-0.0876 ***	-0.0876 ***	-0.0871 ***
	(0.0090)	(0.0090)	(0.0090)	(0.0089)	(0.0091)	(0.0091)
Age$_{it}$	0.0349 ***	0.0354 ***	0.0334 ***	0.0343 ***	0.0371 ***	0.0374 ***
	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0029)	(0.0029)
Age$_{it}^2$	-0.0003 ***	-0.0004 ***	-0.0003 ***	-0.0003 ***	-0.0004 ***	-0.0004 ***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
ln(*nbAuthor$_{it}$*)	0.1855 ***	0.1922 ***	0.1892 ***	0.1966 ***	0.1998 ***	0.2031 ***
	(0.0059)	(0.0059)	(0.0059)	(0.0058)	(0.0059)	(0.0059)
ln(*PublicfundingO$_{it}$*)	0.0206 ***	0.0218 ***	0.0198 ***	0.0213 ***	0.0240 ***	0.0239 ***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
ln(*PrivatefundingO$_{it}$*)	0.0077 ***	0.0078 ***	0.0086 ***	0.0087 ***	0.0041 ***	0.0060 ***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0007)
ln(*NFPfundingO$_{it}$*)	0.0050 ***	0.0065 ***	0.0058 ***	0.0071 ***	0.0064 ***	0.0073 ***
	(0.0008)	(0.0007)	(0.0008)	(0.0007)	(0.0007)	(0.0007)
ColT90$_{it}$	0.1836 ***					
	(0.0068)					
ColT95$_{it}$		0.1863 ***				
		(0.0078)				
ColPub90$_{it}$			0.1820 ***			
			(0.0067)			
ColPub95$_{it}$				0.1841 ***		
				(0.0076)		
ColPriv90$_{it}$					0.1561 ***	
					(0.0073)	
ColPriv95$_{it}$						0.1486 ***
						(0.0084)
ColT90$_{it}$*ln(NFPfundingO$_{it}$)	0.0040 ***					
	(0.0010)					
ColT95$_{it}$*ln(NFPfundingO$_{it}$)		0.0030 ***				
		(0.0011)				
ColPub90$_{it}$*ln(NFPfundingO$_{it}$)			0.0042 ***			
			(0.0010)			
ColPub95$_{it}$*ln(NFPfundingO$_{it}$)				0.0030 ***		
				(0.0011)		
ColPriv90$_{it}$*ln(NFPfundingO$_{it}$)					0.0052 ***	
					(0.0010)	
ColPriv95$_{it}$*ln(NFPfundingO$_{it}$)						0.0050 ***
						(0.0010)
Constant	-0.1445 **	-0.1533 **	-0.1033	-0.1289 *	-0.2370 ***	-0.2346 ***
	(0.0729)	(0.0729)	(0.0726)	(0.0726)	(0.0733)	(0.0735)
Number of observations	35684	35684	35684	35684	35684	35684
Number of scientists	5772	5772	5772	5772	5772	5772
$\chi 2$	7567.56	7403.18	7725.21	7520.69	7116.95	6846.46
sigma	0.3735	0.3744	0.3735	0.3747	0.3746	0.3757
rho	0.2435	0.2400	0.2375	0.2319	0.2457	0.2457
R^2 within groups	0.0634	0.0586	0.0632	0.0569	0.0571	0.0513
R^2 overall	0.2782	0.2755	0.2821	0.2783	0.2678	0.2634
R^2 between groups	0.4668	0.4669	0.4732	0.4727	0.4578	0.4543

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

APPENDIX Y – VARIABLE DESCRIPTION OF CHAPTER 5

Variable name	*Variable description*
dIndChair	Dummy variables taking the value 1 if a scientist has a research chair awarded by industry (industrial chair)
dGCChair	Dummy variables taking the value 1 if a scientist has a research chair awarded by Canadian funding agencies (NSERC, SSHRC, and CIHR)
dCRC	Dummy variables taking the value 1 if a scientist has a Canada research chair
dIndGCChair	Dummy variables taking the value 1 if *dIndChair* or *dGCChair* are equal to 1
dChair	Dummy variables taking the value 1 if any of *dIndChair*, *dGCChair*, or *dCRC* is equal to 1
ln(PublicfundingO)	Natural logarithm of the three-year average of public sector funding for the purpose of operational costs and direct expenditure of research
ln(PrivatefundingO)	Natural logarithm of the three-year average of private sector funding for the purpose of operational costs and direct expenditure of research
ln(NFPfundingO)	Natural logarithm of three-year average of funding from not-for-profit institutions (NFP) for the purpose of operational costs and direct expenditure of research
ln(nbArticle)	Natural logarithm of the yearly number of articles
PubORank	Normalized rank of researcher in the research division in terms of three-year average of funding for the purpose of operational costs and direct expenditure of research
PublRank	Normalized rank of researcher in the research division in terms of three-year average of number of articles
nbScientistUni	Number of scientists in the university and division[45] of researcher
Age	Age of a scientist
dFemale	Dummy variable taking the value 1 if the scientist is a woman and 0 otherwise
ULaval, dUdeM, dUQ_UBishop, dUConcordia, dUMcGill dUSherbrooke, dUQAM_ETS_INRS	Dummy variables indicating the university affiliation of researcher
dMedical, dScience, dBusinessManagement, dEducation, dEngineering, dHealthScience, dHumanities, dNonhealth, and dSocialScience	Dummy variables indicating the research division of researcher
d2000, d2001, …., d2012	Dummy variables indicating the year

[45] There are 9 divisions of Basic Medical Sciences, Business & Management, Education, Engineering, Health Sciences, Humanities, Non-Health Professional, Sciences, and Social Sciences

APPENDIX Z – SUMMARY STATISTICS OF CHAPTER 5

	Average	*Standard Deviation*	*Minimum*	*Maximum*
nbArticles	1.3881	2.8361	0	85
PublicfundingO	64,890.39	151,411.80	0	10,100,000
PrivatefundingO	10,885.52	68,773.97	0	6,257,323
NFPfundingO	11,070.96	84,990.96	0	8,404,625
Age	52.6287	9.9491	1	94
dFemale	0.2829	0.4504	0	1
nbAuthor	0.8476	0.8295	0	7.6327
PubORank	0.5663	0.2576	0.0005	1
PublRank	0.7442	0.1833	0.3206	1
nbScientistUni	5.5921	0.8049	0.6931	7.3803
dCRC	0.0391	0.1938	0	1
dIndGCChair	0.0206	0.1421	0	1
dChair	0.0569	0.2317	0	1

The variables are not summarized in logarithmic scale and they are raw amount. In some disciplines of Physics, there are many scientists involved in one project and therefore, the maximum for the number of authors is high. (Number of Observations: 80775)

APPENDIX AA – CORRELATION TABLE OF CHAPTER 5

		1	2	3	4	5	6	7	8	9	10	11	12	13
ln(nbArticles)	**1**	1												
ln(PublicfundingO)	**2**	0.3543	1											
ln(PrivatefundingO)	**3**	0.3207	0.2120	1										
ln(NFPfundingO)	**4**	0.3093	0.2605	0.2843	1									
Age	**5**	-0.1415	-0.2794	-0.0742	-0.1140	1								
dFemale	**6**	-0.1100	-0.0014	-0.1311	-0.0329	-0.1298	1							
ln(nbAuthor)	**7**	0.7127	0.3502	0.3285	0.3262	-0.1263	-0.1013	1						
PubORank	**8**	0.3502	0.7036	0.2099	0.2633	-0.1371	-0.0052	0.3375	1					
PublRank	**9**	0.3382	0.2770	0.0919	0.1300	-0.1196	0.0508	0.3204	0.4026	1				
ln(nbScientistUni)	**10**	0.1545	-0.0152	0.1216	0.1157	0.0617	-0.0767	0.2442	0.0454	-0.0902	1			
dCRC	**11**	0.1554	0.1590	0.1032	0.1112	-0.0743	-0.0385	0.1090	0.2258	0.1052	-0.0223	1		
dIndGCChair	**12**	0.0800	0.0872	0.2112	0.1017	0.0095	-0.0486	0.0529	0.1082	0.0044	-0.0055	0.0724	1	
dChair	**13**	0.1700	0.1780	0.1959	0.1432	-0.0543	-0.0572	0.1171	0.2426	0.0866	-0.0243	0.8211	0.5906	1

Significant at 1% level (No. observations: 80775)

APPENDIX BB – FIRST STAGE REGRESSION OF TABLE 5.2

	xtreg1	*xtreg2*	*xtreg3*	*xtreg4*	*xtreg5*
$ln(PrivatefundingO_{it})$	-0.0048	-0.0049 *	-0.0049 *	-0.0034	-0.0034
	(0.0029)	(0.0029)	(0.0029)	(0.0030)	(0.0030)
$ln(NFPfundingO_{it})$	0.0211 ***	0.0214 ***	0.0214 ***	0.0214 ***	0.0214 ***
	(0.0026)	(0.0026)	(0.0026)	(0.0026)	(0.0026)
$dFemale_i$	0.2437 ***	0.2396 ***	0.2374 ***	0.2401 ***	0.2391 ***
	(0.0384)	(0.0380)	(0.0385)	(0.0379)	(0.0384)
Age_{it}	0.1395 ***	0.1364 ***	0.1364 ***	0.1361 ***	0.1361 ***
	(0.0105)	(0.0105)	(0.0105)	(0.0104)	(0.0104)
Age_{it}^2	-0.0019 ***	-0.0018 ***	-0.0018 ***	-0.0018 ***	-0.0018 ***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$ln(nbAuthor_{it})$	0.1965 ***	0.1956 ***	0.1956 ***	0.1949 ***	0.1949 ***
	(0.0214)	(0.0214)	(0.0214)	(0.0214)	(0.0214)
$dCRC_{it}$		-0.2791 ***	-0.2945 ***	-0.1849 *	-0.1926 *
		(0.0873)	(0.0974)	(0.0970)	(0.1078)
$dCRC_{it}*dFemale_i$			0.0744		0.0343
			(0.2116)		(0.2118)
$dCRC_{it}*ln(PrivatefundingO_{it})$				-0.0255 **	-0.0253 **
				(0.0115)	(0.0115)
$PubORank_{it-1}$	10.0707 ***	10.1046 ***	10.1057 ***	10.1080 ***	10.1084 ***
	(0.0485)	(0.0488)	(0.0488)	(0.0488)	(0.0488)
$PublRank_{it-1}$	1.5967 ***	1.6057 ***	1.6057 ***	1.6046 ***	1.6047 ***
	(0.0957)	(0.0958)	(0.0958)	(0.0957)	(0.0957)
$ln(nbScientistUni_{it-1})$	-0.4284 ***	-0.4339 ***	-0.4340 ***	-0.4349 ***	-0.4350 ***
	(0.0259)	(0.0257)	(0.0257)	(0.0256)	(0.0256)
Constant	0.7895 **	0.8879 ***	0.8894 ***	0.8982 ***	0.8988 ***
	(0.3235)	(0.3225)	(0.3223)	(0.3219)	(0.3219)
Number of observations	80775	80775	80775	80775	80775
Number of scientists	7651	7651	7651	7651	7651
χ^2	75797 ***	76288 ***	76318 ***	76438 ***	76450 ***
R^2 within groups	0.2082	0.2080	0.2080	0.2079	0.2079
R^2 overall	0.5459	0.5504	0.5506	0.5511	0.5512
R^2 between groups	0.7486	0.7548	0.7551	0.756	0.7562

Notes: *, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant - The reported numbers for R2 are approximation because they are not reported in the first stage but we obtain them from re-doing the first stage independently.

APPENDIX CC – FIRST STAGE REGRESSION OF TABLE 5.3

	xtreg1	*xtreg2*	*xtreg3*	*xtreg4*	*xtreg5*
ln(PrivatefundingO$_{it}$)	-0.0048	-0.0037	-0.0037	-0.0039	-0.0038
	(0.0029)	(0.0029)	(0.0029)	(0.0030)	(0.0030)
ln(NFPfundingO$_{it}$)	0.0211 ***	0.0215 ***	0.0215 ***	0.0215 ***	0.0215 ***
	(0.0026)	(0.0026)	(0.0026)	(0.0026)	(0.0026)
dFemale$_i$	0.2437 ***	0.2440 ***	0.2459 ***	0.2439 ***	0.2458 ***
	(0.0384)	(0.0384)	(0.0386)	(0.0383)	(0.0386)
Age$_{it}$	0.1395 ***	0.1395 ***	0.1396 ***	0.1394 ***	0.1395 ***
	(0.0105)	(0.0105)	(0.0105)	(0.0105)	(0.0105)
Age$_{it}^2$	-0.0019 ***	-0.0019 ***	-0.0019 ***	-0.0019 ***	-0.0019 ***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
ln(nbAuthor$_{it}$)	0.1965 ***	0.1952 ***	0.1952 ***	0.1952 ***	0.1952 ***
	(0.0214)	(0.0214)	(0.0214)	(0.0214)	(0.0214)
dIndGCChair$_{it}$		-0.4995 ***	-0.4778 ***	-0.5273 ***	-0.5052 ***
		(0.1244)	(0.1328)	(0.1660)	(0.1727)
dIndGCChair$_{it}$*dFemale$_i$			-0.1650		-0.1641
			(0.3539)		(0.3537)
dIndGCChair$_{it}$*ln(PrivatefundingO$_{it}$)				0.0037	0.0036
				(0.0147)	(0.0147)
PubORank$_{it-1}$	10.0707 ***	10.0824 ***	10.0823 ***	10.0829 ***	10.0828 ***
	(0.0485)	(0.0486)	(0.0486)	(0.0486)	(0.0486)
PublRank$_{it-1}$	1.5967 ***	1.6056 ***	1.6051 ***	1.6051 ***	1.6046 ***
	(0.0957)	(0.0957)	(0.0957)	(0.0957)	(0.0957)
ln(nbScientistUni$_{it-1}$)	-0.4284 ***	-0.4275 ***	-0.4275 ***	-0.4275 ***	-0.4275 ***
	(0.0259)	(0.0259)	(0.0259)	(0.0259)	(0.0259)
Constant	0.7895 **	0.7590 **	0.7584 **	0.7618 **	0.7612 **
	(0.3235)	(0.3235)	(0.3235)	(0.3235)	(0.3235)
Number of observations	80775	80775	80775	80775	80775
Number of scientists	7651	7651	7651	7651	7651
χ^2	75797 ***	75835 ***	75838 ***	75859 ***	75862 ***
R^2 within groups	0.2082	0.2083	0.2083	0.2083	0.2083
R^2 overall	0.5459	0.5462	0.5463	0.5463	0.5464
R^2 between groups	0.7486	0.7489	0.749	0.7491	0.7491

Notes: *, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant - The reported numbers for R2 are approximation because they are not reported in the first stage but we obtain them from re-doing the first stage independently.

APPENDIX DD – FIRST STAGE REGRESSION OF TABLE 5.4

	xtreg1	*xtreg2*	*xtreg3*	*xtreg4*	*xtreg5*
ln(PrivatefundingO$_{it}$)	-0.0048	-0.0042	-0.0042	-0.0024	-0.0024
	(0.0029)	(0.0029)	(0.0029)	(0.0031)	(0.0031)
ln(NFPfundingO$_{it}$)	0.0211 ***	0.0217 ***	0.0217 ***	0.0217 ***	0.0217 ***
	(0.0026)	(0.0026)	(0.0026)	(0.0026)	(0.0026)
dFemale$_i$	0.2437 ***	0.2391 ***	0.2369 ***	0.2400 ***	0.2390 ***
	(0.0384)	(0.0380)	(0.0388)	(0.0380)	(0.0387)
Age$_{it}$	0.1395 ***	0.1363 ***	0.1363 ***	0.1364 ***	0.1363 ***
	(0.0105)	(0.0105)	(0.0105)	(0.0104)	(0.0104)
Age$_{it}^2$	-0.0019 ***	-0.0018 ***	-0.0018 ***	-0.0018 ***	-0.0018 ***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
ln(nbAuthor$_{it}$)	0.1965 ***	0.1950 ***	0.1949 ***	0.1942 ***	0.1942 ***
	(0.0214)	(0.0214)	(0.0214)	(0.0214)	(0.0214)
dChair$_{it}$		-0.3415 ***	-0.3518 ***	-0.2579 ***	-0.2632 ***
		(0.0745)	(0.0820)	(0.0857)	(0.0937)
dChair5$_{it}$*dFemale$_i$			0.0547		0.0255
			(0.1850)		(0.1853)
dChair5$_{it}$*ln(PrivatefundingO$_{it}$)				-0.0181 **	-0.0180 **
				(0.0092)	(0.0093)
PubORank$_{it-1}$	10.0707 ***	10.1136 ***	10.1144 ***	10.1165 ***	10.1169 ***
	(0.0485)	(0.0489)	(0.0489)	(0.0489)	(0.0489)
PublRank$_{it-1}$	1.5967 ***	1.6149 ***	1.6150 ***	1.6150 ***	1.6150 ***
	(0.0957)	(0.0958)	(0.0958)	(0.0958)	(0.0958)
ln(nbScientistUni$_{it-1}$)	-0.4284 ***	-0.4346 ***	-0.4346 ***	-0.4350 ***	-0.4351 ***
	(0.0259)	(0.0257)	(0.0257)	(0.0257)	(0.0257)
Constant	0.7895 ***	0.8798 ***	0.8810 ***	0.8733 ***	0.8740 ***
	(0.3235)	(0.3223)	(0.3222)	(0.3220)	(0.3220)
Number of observations	80775	80775	80775	80775	80775
Number of scientists	7651	7651	7651	7651	7651
χ^2	75797 ***	76248 ***	76270 ***	76338 ***	76349 ***
R^2 within groups	0.2082	0.2082	0.2083	0.2082	0.2082
R^2 overall	0.5459	0.5502	0.5504	0.5506	0.5508
R^2 between groups	0.7486	0.7545	0.7548	0.7552	0.7554

Notes: *, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant - The reported numbers for R2 are approximation because they are not reported in the first stage but we obtain them from re-doing the first stage independently.

APPENDIX EE – FIRST STAGE REGRESSION OF TABLE 5.5

	xtreg1	*xtreg2*	*xtreg3*	*xtreg4*	*xtreg5*
ln(PrivatefundingO$_{it}$)	0.0384 ***	0.0387 ***	0.0386 ***	0.0551 ***	0.0546 ***
	(0.0050)	(0.0050)	(0.0050)	(0.0065)	(0.0065)
ln(NFPfundingO$_{it}$)	0.0306 ***	0.0306 ***	0.0306 ***	0.0307 ***	0.0307 ***
	(0.0048)	(0.0048)	(0.0048)	(0.0048)	(0.0048)
dFemale$_i$	-0.1671	-0.1820	-0.4000 **	-0.1792	-0.3540 *
	(0.1283)	(0.1270)	(0.1896)	(0.1269)	(0.1898)
Age$_{it}$	0.0905 ***	0.0951 ***	0.0966 ***	0.0951 ***	0.0964 ***
	(0.0278)	(0.0279)	(0.0279)	(0.0278)	(0.0279)
Age$_{it}^2$	-0.0011 ***	-0.0011 ***	-0.0011 ***	-0.0011 ***	-0.0011 ***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
ln(nbAuthor$_{it}$)	0.1554 ***	0.1569 ***	0.1568 ***	0.1588 ***	0.1586 ***
	(0.0513)	(0.0513)	(0.0513)	(0.0512)	(0.0512)
dCRC$_{it}$		0.2235 **	0.1596	0.3833 ***	0.3280 ***
		(0.0991)	(0.1073)	(0.1070)	(0.1159)
dCRC$_{it}$*dFemale$_i$			0.3792		0.3040
			(0.2447)		(0.2454)
dCRC$_{it}$*ln(PrivatefundingO$_{it}$)				-0.0378 ***	-0.0368 ***
				(0.0096)	(0.0096)
PubORank$_{it-1}$	6.4595 ***	6.4423 ***	6.4395 ***	6.4371 ***	6.4348 ***
	(0.1224)	(0.1229)	(0.1229)	(0.1227)	(0.1227)
PublRank$_{it-1}$	0.8820 ***	0.8443 ***	0.8515 ***	0.8452 ***	0.8511 ***
	(0.2087)	(0.2089)	(0.2089)	(0.2086)	(0.2087)
ln(nbScientistUni$_{it-1}$)	-0.2852 ***	-0.2669 ***	-0.2730 ***	-0.2722 ***	-0.2769 ***
	(0.0700)	(0.0697)	(0.0698)	(0.0697)	(0.0698)
Constant	4.5170 ***	4.2376 ***	4.2679 ***	4.1752 ***	4.2009 ***
	(0.8221)	(0.8278)	(0.8279)	(0.8272)	(0.8274)
Number of observations	6393	6393	6393	6393	6393
Number of scientists	586	586	586	586	586
χ^2	4260 ***	4288 ***	4292 ***	4313 ***	4315 ***
R^2 within groups	0.2527	0.2519	0.2518	0.2521	0.2521
R^2 overall	0.5038	0.5084	0.5087	0.5092	0.5093
R^2 between groups	0.6404	0.6471	0.6475	0.6483	0.6484

Notes: *, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant - The reported numbers for R2 are approximation because they are not reported in the first stage but we obtain them from re-doing the first stage independently.

APPENDIX FF – FIRST STAGE REGRESSION OF TABLE 5.6

	xtreg1	xtreg2	xtreg3	xtreg4	xtreg5
ln(PrivatefundingO$_{it}$)	0.0457 ***	0.0481 ***	0.0481 ***	-0.0108	-0.0106
	(0.0085)	(0.0086)	(0.0086)	(0.0120)	(0.0120)
ln(NFPfundingO$_{it}$)	0.0304 ***	0.0311 ***	0.0313 ***	0.0323 ***	0.0324 ***
	(0.0076)	(0.0076)	(0.0076)	(0.0076)	(0.0076)
dFemale$_i$	0.0215	0.0476	0.3081	0.0150	0.2072
	(0.2363)	(0.2366)	(0.3577)	(0.2355)	(0.3562)
Age$_{it}$	0.3513 ***	0.3537 ***	0.3552 ***	0.3461 ***	0.3473 ***
	(0.0463)	(0.0463)	(0.0464)	(0.0460)	(0.0461)
Age$_{it}^2$	-0.0037 ***	-0.0037 ***	-0.0037 ***	-0.0036 ***	-0.0036 ***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
ln(nbAuthor$_{it}$)	0.1862 **	0.1735 **	0.1768 **	0.1511 *	0.1536 *
	(0.0803)	(0.0805)	(0.0806)	(0.0800)	(0.0801)
dIndGCChair$_{it}$		-0.3138 **	-0.2690	-0.9872 ***	-0.9519 ***
		(0.1646)	(0.1712)	(0.1894)	(0.1961)
dIndGCChair$_{it}$*dFemale$_i$			-0.4566		-0.3369
			(0.4696)		(0.4675)
dIndGCChair$_{it}$*ln(PrivatefundingO$_{it}$)				0.1119 ***	0.1115 ***
				(0.0158)	(0.0158)
PubORank$_{it-1}$	6.8202 ***	6.8098 ***	6.8054 ***	6.7298 ***	6.7268 ***
	(0.1870)	(0.1870)	(0.1870)	(0.1859)	(0.1860)
PublRank$_{it-1}$	1.3377 ***	1.3282 ***	1.3256 ***	1.2888 ***	1.2871 ***
	(0.3000)	(0.3000)	(0.3001)	(0.2979)	(0.2980)
ln(nbScientistUni$_{it-1}$)	-0.4670 ***	-0.4341 **	-0.4282 **	-0.4165 **	-0.4122 **
	(0.1786)	(0.1794)	(0.1798)	(0.1785)	(0.1789)
Constant	-1.6525	-1.8467	-1.9517	-1.3026	-1.3824
	(1.6226)	(1.6254)	(1.6305)	(1.6175)	(1.6228)
Number of observations	3234	3234	3234	3234	3234
Number of scientists	288	288	288	288	288
χ^2	2879 ***	2885 ***	2884 ***	2977 ***	2976 ***
R^2 within groups	0.2396	0.2396	0.2397	0.24	0.2401
R^2 overall	0.4085	0.4083	0.4083	0.408	0.4079
R^2 between groups	0.5395	0.5392	0.539	0.5383	0.5381

Notes: *, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant - The reported numbers for R2 are approximation because they are not reported in the first stage but we obtain them from re-doing the first stage independently.

APPENDIX GG – FIRST STAGE REGRESSION OF TABLE 5.7

	xtreg1	*xtreg2*	*xtreg3*	*xtreg4*	*xtreg5*
ln(PrivatefundingO$_{it}$)	0.0281 ***	0.0281 ***	0.0281 ***	0.0135 **	0.0133 **
	(0.0047)	(0.0047)	(0.0047)	(0.0064)	(0.0064)
ln(NFPfundingO$_{it}$)	0.0266 ***	0.0266 ***	0.0266 ***	0.0265 ***	0.0266 ***
	(0.0044)	(0.0044)	(0.0044)	(0.0044)	(0.0044)
dFemale$_i$	-0.0492	-0.0464	-0.1461	-0.0473	-0.1738
	(0.1141)	(0.1134)	(0.1664)	(0.1134)	(0.1667)
Age$_{it}$	0.1203 ***	0.1194 ***	0.1202 ***	0.1172 ***	0.1181 ***
	(0.0247)	(0.0246)	(0.0247)	(0.0246)	(0.0247)
Age$_{it}^2$	-0.0014 ***	-0.0014 ***	-0.0014 ***	-0.0014 ***	-0.0014 ***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
ln(nbAuthor$_{it}$)	0.1686 ***	0.1673 ***	0.1675 ***	0.1640 ***	0.1642 ***
	(0.0453)	(0.0453)	(0.0453)	(0.0453)	(0.0453)
dChair$_{it}$		-0.0492	-0.0771	-0.1837 *	-0.2219 **
		(0.0875)	(0.0939)	(0.0963)	(0.1031)
dChair5$_{it}$*dFemale$_i$			0.1819		0.2309
			(0.2223)		(0.2228)
dChair5$_{it}$*ln(PrivatefundingO$_{it}$)				0.0292 ***	0.0298 ***
				(0.0087)	(0.0088)
PubORank$_{it-1}$	7.0954 ***	7.1057 ***	7.1028 ***	7.1086 ***	7.1050 ***
	(0.1067)	(0.1071)	(0.1072)	(0.1071)	(0.1072)
PublRank$_{it-1}$	1.1549 ***	1.1578 ***	1.1579 ***	1.1592 ***	1.1593 ***
	(0.1783)	(0.1784)	(0.1784)	(0.1783)	(0.1783)
ln(nbScientistUni$_{it-1}$)	-0.1736 ***	-0.1772 ***	-0.1794 ***	-0.1777 ***	-0.1806 ***
	(0.0653)	(0.0652)	(0.0653)	(0.0652)	(0.0653)
Constant	2.5660 ***	2.6143 ***	2.6237 ***	2.7596 ***	2.7745 ***
	(0.7358)	(0.7375)	(0.7377)	(0.7387)	(0.7389)
Number of observations	9092	9092	9092	9092	9092
Number of scientists	836	836	836	836	836
χ^2	7043 ***	7064 ***	7063 ***	7080 ***	7081 ***
R^2 within groups	0.2514	0.2513	0.2513	0.2516	0.2516
R^2 overall	0.4828	0.4878	0.4878	0.4876	0.4877
R^2 between groups	0.6317	0.6393	0.6394	0.6388	0.6388

Notes: *, **, and *** show the significance level at 0.1, 0.05, and 0.01 respectively - Year dummies, research division dummies, and university dummies are significant - The reported numbers for R2 are approximation because they are not reported in the first stage but we obtain them from re-doing the first stage independently.